Université de Montréal

en cotutelle avec

Université Paris-Diderot

Structures linéaires dans les ensembles à faible densité

par

Kevin Henriot

Département de mathématiques et de statistique

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.) en Mathématiques

Avril 2014

Université de Montréal

Faculté des Etudes Supérieures

Université Paris-Diderot

École Doctorale de Sciences Mathématiques de Paris Centre

Cette thèse intitulée

Structures linéaires dans les ensembles à faible densité

---

a été évaluée par le jury suivant :

Régis de la Bretèche	Directeur de recherche
Andrew Granville	Directeur de recherche
Hamed Hatami	Membre du jury
Dimitris Koukoulopoulos	Président-rapporteur
Tom Sanders	Examinateur externe
	Représentant du doyen de la FESP

Thèse acceptée le :
23 avril 2014

Université Paris-Diderot

Université de Montréal

Ecole Doctorale de Sciences Mathématiques de Paris Centre

Thèse de doctorat

Discipline : Mathématiques

présentée par

Kevin Henriot

Structures linéaires dans les ensembles à faible densité

---

dirigée par Régis de la Bretèche et Andrew Granville

a été soutenue le

devant le jury suivant :

Régis de la Bretèche	Université Paris-Diderot	Directeur
Andrew Granville	Université de Montréal	Directeur
Hamed Hatami	McGill University	Examinateur
Dimitris Koukoulopoulos	Université de Montréal	Examinateur
Tom Sanders	University of Oxford	Rapporteur

au vu des rapports de :

Izabella Łaba	University of British Columbia
Tom Sanders	University of Oxford

Résumé

Nous présentons trois résultats en combinatoire additive, un domaine récent à la croisée de la combinatoire, l'analyse harmonique et la théorie analytique des nombres. Le thème unificateur de notre thèse est la détection de structures additives dans les ensembles arithmétiques à faible densité, avec un intérêt particulier pour les aspects quantitatifs. Notre première contribution est une estimation de densité améliorée pour le problème, initié entre autres par Bourgain, de trouver une longue progression arithmétique dans un ensemble somme triple. Notre deuxième résultat consiste en une généralisation des bornes de Sanders pour le théorème de Roth, du cas d'un ensemble dense dans les entiers à celui d'un ensemble à faible croissance additive dans un groupe abélien arbitraire. Finalement, nous étendons les meilleures bornes quantitatives connues pour le théorème de Roth dans les premiers, à tous les systèmes d'équations linéaires invariants par translation et de complexité un.

Mots-clés : combinatoire additive, progressions arithmétiques, ensembles sommes, théorème de Freiman-Ruzsa, théorème de Roth, théorème de Green-Tao, équations linéaires dans les nombres premiers.

Summary

We present three results in additive combinatorics, a recent field at the interface of combinatorics, harmonic analysis and analytic number theory. The unifying theme in our thesis is the detection of additive structure in arithmetic sets of low density, with an emphasis on quantitative aspects. Our first contribution is an improved density estimate for the problem, initiated by Bourgain and others, of finding a long arithmetic progression in a triple sumset. Our second result is a generalization of Sanders' bounds for Roth's theorem from the dense setting, to the setting of small doubling in an arbitrary abelian group. Finally, we extend the best known quantitative results for Roth's theorem in the primes, to all translation-invariant systems of equations of complexity one.

Keywords : additive combinatorics, arithmetic progressions in sumsets, Freiman-Ruzsa theorem, Roth's theorem, Green-Tao theorem, linear equations in primes.

Table des matières

Remerciements

Nous remercions nos directeurs de recherche, Régis de la Bretèche et Andrew Granville, pour leur soutien et des encouragements à tous les stades de cette thèse, ainsi que pour de nombreux conseils avisés sur l'écriture et la communication mathématique.

Nous remercions Tom Sanders pour de nombreux encouragements sur des versions préliminaires des résultats contenus dans cette thèse. Nous remercions les deux rapporteurs externes d'avoir accepté la lourde tâche d'écrire un rapport de thèse.

Nous remercions toute l'équipe administrative du Département de Mathématiques, et Anne-Marie Dupuis en particulier, pour nous avoir guidé au travers des règles obscures de l'Université de Montréal.

Nous remercions les institutions qui ont soutenu financièrement ce travail : l'École Normale Supérieure de Lyon, l'Université Paris-Diderot et la Chaire de Recherche Canadienne d'Andrew Granville.

Nous remercions nos amis dans le groupe des thésards en théorie des nombres, pour les bons moments passés ensemble, et pour beaucoup de discussions mathématiques stimulantes : Farzad Aryan, Mohammad Bardestani, Crystel Bujold, Dimitri Dias, Daniel Fiorilli, Tristan Freiberg, Oleksiy Klurman et Marzieh Mehdizadeh. Nous remercions aussi tous nos autres amis au Département de Mathématique, pour avoir créé l'atmosphère unique de cet endroit.

Nous remercions notre famille : Agnès et Patrick, Christian et Feng Yi ; pour leur amour et leur soutien inconditionnels au cours du long chemin qui a mené à cette thèse. Enfin, les mots ne suffiraient pas à décrire combien nous devons à notre partenaire Golnaz, dans le travail comme dans la vie.

Chapitre I. Introduction

1. Survol des résultats

Le théorème de Roth [69] est considéré de nos jours comme un résultat pionnier de la combinatoire additive, et son énoncé est très simple : tout sous-ensemble des entiers de densité asymptotique strictement positive contient une progression arithmétique à trois termes non triviale. Cela répondait à une version faible d'une conjecture faite par Erdös et Turan [16] en 1936, qui prédit que tout ensemble $\mathcal{A} \subset \mathbb{N}$ tel que

$$\sum_{a \in \mathcal{A}} \frac{1}{a} = \infty \tag{1.1}$$

contient une progression arithmétique à k termes, pour tout $k \geqslant 3$. Pour quantifier ce type de résultats, nous considérons dorénavant un sous-ensemble A de $[N]$, où N est un entier qui tend vers l'infini, et nous appelons $\alpha = |A|/N$ la densité de A. La méthode de Roth [69] permet en réalité de détecter des progressions arithmétiques dans A pour une densité descendant jusqu'à $C(\log\log N)^{-1}$, et les travaux subséquents de Heath-Brown [48] et Szemerédi [96] ont montré que le résultat reste valable pour une densité $(\log N)^{-c}$, où $c > 0$ est une petite constante. De nouvelles méthodes importantes ont été introduites par Bourgain [5] dans sa preuve que tout exposant $c < 1/2$ est admissible, et des progrès successifs ont ensuite été accomplis par Bourgain [6] et Sanders [82], jusqu'à la récente percée de ce dernier.

THÉORÈME (Sanders [81]). *Soit A un sous-ensemble de $[N]$ de densité au moins*

$$C(\log N)^{-1}(\log\log N)^{5}.$$

Alors A contient une progression arithmétique à trois termes non triviale.

Par sommation partielle, on peut vérifier que les ensembles $\mathcal{A} \subset \mathbb{N}$ satisfaisant (1.1) ont une densité au moins égale à $(\log N)^{-1}(\log\log N)^{-1-\varepsilon}$ dans les N premiers entiers, et donc le résultat ci-dessus réussit presque à établir la conjecture d'Erdös-Turan (pour $k = 3$) ; il semble toutefois que de nouvelles idées sont nécessaires pour dépasser la « barrière logarithmique ». Le résultat de Sanders a été étendu par la suite par Bloom [2], qui a montré que toute équation invariante par translation en $s \geqslant 3$ variables, comme par exemple $x_1 + \cdots + x_{s-1} = (s-1)x_s$, est résoluble non trivalement dans un sous-ensemble de $[N]$ de densité[1] $\alpha \gtrsim (\log N)^{-(s-2)}$.

Un problème voisin est de détecter certaines structures additives dans l'ensemble somme

$$A + A = \{a + a' : (a, a') \in A^2\}.$$

Un résultat étonnant de Bourgain [4] dans cette veine dit que lorsque A a pour densité $\alpha \gtrsim (\log N)^{-1/2}$, l'ensemble somme[2] $A + A$ contient une progression arithmétique de longueur au moins

$$\exp\left[c(\alpha^2 \log N)^{1/3}\right].$$

L'exposant $1/3$ a par la suite été amélioré à $1/2$ par Green [29], et le domaine de densité admissible à $\alpha \gtrsim (\log N)^{-1}$ par Croot, Laba and Sisask [9] : remarquons la similarité avec le domaine de densité pour le théorème de Roth. D'un autre côté,

[1] Nous écrivons $X \gtrsim Y$ pour une condition de la forme $X \geqslant CY(\log Y)^C$ avec une constante $C > 0$ non spécifiée.

[2] Les résultats que nous citons s'appliquent aussi aux ensembles sommes asymmétriques de la forme $A + B$, mais nous nous restreignons au cas symmétrique pour simplifier l'exposition.

le travail de Sanders [78] qui améliore des résultats précédents [18, 29] permet de trouver des progressions arithmétiques de longueur $N^{c\alpha^{1+o(1)}}$ dans l'ensemble somme triple $A+A+A$, quoique uniquement dans le domaine $\alpha \gtrsim (\log N)^{-1/2}$. Notre premier résultat abaisse cette densité à $(\log N)^{-2}$, ce qui est à nouveau comparable avec les bornes connues pour le théorème de Roth.

THÉORÈME 1. *Soit A un sous-ensemble de $[N]$ de densité α. Alors $A+A+A$ contient une progression arithmétique de longueur au moins*

$$\exp\left[c\alpha^{1/4}(\log \alpha^{-1})^{-7/2}(\log N)^{1/2}\right] \quad \textit{pourvu que} \quad \alpha \geqslant C(\log N)^{-2}(\log\log N)^{14}.$$

Un autre résultat fondamental de combinatoire additive est le théorème de Freiman-Ruzsa [17, 77], qui décrit la structure approximative des ensembles d'entiers à faible croissance additive. Ce théorème dit que, si A est un ensemble fini d'entiers tel que $|A+A| \leqslant K|A|$ pour un paramètre $K \geqslant 1$, alors A est contenu dans une progression arithmétique généralisée (PAG) $Q = \{n_1u_1 + \cdots + n_du_d : 0 \leqslant n_i \leqslant N_i\}$ (où $u_i \in \mathbb{Z}$ et $N_i \geqslant 1$) telle que $|Q| \leqslant C(K)|A|$ et $d \leqslant C(K)$, où $C(K)$ est une constante dépendant de K. Cet énoncé a été par la suite généralisé à un groupe abélien arbitraire par Green and Ruzsa [32], qui ont adapté légèrement la structure recherchée. On ne peut obtenir une meilleure dépendence que $C(K) = e^{O(K)}$ dans le théorème de Freiman-Ruzsa, ce qui a motivé la conjecture de Freiman-Ruzsa polynomiale [63, 84], laquelle prédit qu'un ensemble d'entiers A tel que $|A+A| \leqslant K|A|$ possède une intersection de taille au moins $|A|/f(K)$ avec une PAG de taille au plus $f(K)|A|$ et de dimension au plus $\log f(K)$, où $f(K) = K^{O(1)}$. Les résultats de ce type ont de nombreuses applications [38, 63, 84, 90], et par conséquent améliorer la borne $f(K)$ est un problème ouvert majeur en combinatoire additive. Les premières bornes efficaces vers la conjecture de Freiman-Ruzsa polynomiale ont été obtenues par Chang [7], et des progrès majeurs ont été accomplis plus récemment par Schoen [88] et Sanders [83, 84], culminant avec la preuve par ce dernier que $f(K) = \exp[(\log K)^{3+o(1)}]$ est admissible.

Dans notre second travail, nous ne faisons pas de progrès sur ce problème important, mais nous posons à la place une question voisine : peut-on trouver une structure additive exacte, en l'occurence une progression arithmétique à trois termes, au lieu d'une structure additive approximative, dans un ensemble à faible doublement ? Dans le cadre général d'un groupe abélien, on peut répondre qualitativement à cette question à l'aide des techniques de modélisation de Green et Ruzsa [32], mais sur le plan quantitatif le problème est plus délicat. Sanders [80] a examiné la question posée, et a montré que tout sous-ensemble fini A d'un groupe abélien de doublement au plus $(\log |A|)^{1/3-o(1)}$ contient une progression arithmétique non triviale. Nous améliorons aussi ce résultat, et nous obtenons des bornes de la qualité de celles connues pour le théorème de Roth.

THÉORÈME 2. *Soit A un sous-ensemble fini d'un groupe abélien tel que*

$$\frac{|A+A|}{|A|} \leqslant \frac{c \log |A|}{(\log \log |A|)^7}.$$

Alors A contient une progression arithmétique à trois termes dont les termes ne sont pas tous égaux.

La motivation d'origine derrière la conjecture d'Erdös-Turan était que sa résolution viendrait à bout d'un problème ouvert à l'époque : le fait que les premiers contiennent des progressions arithmétiques arbitrairement longues. Il est bien connu que ce problème a été résolu par Green and Tao [36] en 2004, et leur preuve montre de plus que le résultat vaut pour tout sous-ensemble des nombres premiers de densité relative strictement positive. La conjecture originale de Erdös-Turan reste cependant ouverte, et Green et Tao ont pu traiter le cas spécifique des nombres premiers en développant un principe de transfert, qui réduit le problème à trouver des progressions arithmétiques arbitrairement longues dans tout sous-ensemble dense des entiers, auquel cas il s'agit précisément du théorème de Szemerédi [95].

Le théorème de Green-Tao s'étend de fait à tout système de forme linéaires entières $\psi = (\psi_1, \dots, \psi_t)$ constitué de formes linéairement indépendantes deux à

deux, et qui plus est *invariant par translation*[3] : on peut déduire de ce théorème que pour tout sous-ensemble des nombres premiers de densité relative strictement positive, on peut toujours trouver une configuration $\psi(x) \in A^t$ non triviale, i.e. à coordonnées distinctes. Dans le contexte de trouver des asymptotiques pour les occurences de configurations linéaires dans les nombres premiers, Green and Tao [**39**] ont défini la notion de *complexité* pour un système de formes linéaires, et la classe des systèmes de complexité 1 peut être décrite comme le domaine d'applicabilité des méthodes d'analyse harmonique classique, alors que les cas de complexité supérieure requièrent des techniques distinctes appartenant à la théorie de l'uniformité d'ordre supérieur [**99**], pour laquelle il existe moins de résultats quantitatifs.

La classe de complexité 1 inclut les progressions arithmétiques à trois termes, mais pas les plus longues, et a été récemment considérée dans le cadre des entiers par Shao [**91**], qui a généralisé les bornes logarithmiques de Bourgain [**5**] pour le théorème de Roth à un système « modèle » de formes linéaires de complexité 1. D'un autre côté, dans le cas des nombres premiers, les résultats quantitatifs obtenus jusqu'à présent ont surtout concerné l'analogue du théorème de Roth : en améliorant le résultat de Green [**30**], Helfgott et de Roton [**50**] ont montré que tout sous-ensemble des nombres premiers jusqu'à N de densité $\gtrsim (\log\log N)^{-1/3}$ contient une progression arithmétique à trois termes. Nous étendons cette borne à toutes les configurations linéaires invariantes par translation et de complexité un.

THÉORÈME 3. *Soient $d, t \geqslant 1$ et $\psi : \mathbb{Z}^d \to \mathbb{Z}^t$ un système de formes linéaires invariant par translation et de complexité un. Soit aussi A un sous-ensemble des premiers jusqu'à N de densité au moins égale à*

$$C(\log\log N)^{-1/24t}.$$

Il existe alors $x \in \mathbb{Z}^d$ tel que $\psi(x) \in A^t$ possède des coordonnées distinctes.

[3] On dit que $\psi : \mathbb{Z}^d \to \mathbb{Z}^t$ est invariant par translation lorsque pour tous $(u_1, \dots, u_t) \in \mathrm{Im}(\psi)$ et $h \in \mathbb{Z}$, on a $(u_1 + h, \dots, u_t + h) \in \mathrm{Im}(\psi)$.

2. Organisation de la thèse

Nous décrivons maintenant la structure de cette thèse. Nos publications ou prépublications sont contenues intégralement dans les Chapitres III–V, et cela nous invite à recommander un ordre de lecture peu conventionnel. En effet, nous suggérons au lecteur de commencer par consulter l'introduction de chacun des Chapitres III–V, où l'on trouve une revue de littérature complète pour les problèmes abordés dans cette thèse. Remarquons que le Théoreme 1 correspond au Chapitre III, le Théorème 2 au Chapitre IV, le Théoreme 3 au Chapitre V.

La prochaine étape recommandée est de parcourir le Chapitre II d'exposition. La première moitié de ce chapitre concerne des notions préliminaires sous-jacentes à tous nos résultats. Nous ne donnons pas de preuves formelles, et nous nous concentrons plutôt sur le contexte dans lequel les techniques que nous utilisons ont été développées, et sur leur signification. Dans la seconde moitié du Chapitre II, nous faisons des commentaires informels sur les résultats de cette thèse. Pour éviter une répétition avec les Chapitres III–V, nous nous bornons à esquisser la preuve de chacun de nos résultats, et à donner quelques informations supplémentaires sur les méthodes globales utilisées. Nous espérons que la lecture du Chapitre II peut aider à comprendre les principales idées présentes dans nos travaux, sans avoir à plonger dans les détails techniques de nos preuves.

Chapitre II. Préliminaires et résumés des travaux

1. Bases : Notation

Nous rappelons un certain nombre de notations qui sont d'usage courant en combinatoire additive, mais peut-être moins en dehors de ce domaine. Nous ajoutons aussi quelques notations originales, qui ne diffèrent guère de la norme.

Lorsque X est un ensemble fini et $f : X \to \mathbb{C}$ est une fonction, nous utilisons tour à tour la notation $\mathbb{E}_X f$ ou $\mathbb{E}_{x\in X} f(x)$ pour désigner la moyenne $|X|^{-1}\sum_{x\in X} f(x)$. Nous écrivons aussi $\mathbb{Z}_N$ pour le groupe cyclique $\mathbb{Z}/N\mathbb{Z}$ lorsque $N \geqslant 1$, bien que dans d'autres contextes cette notation soit réservée pour les N-adiques. Nous posons $[N] = \{1, \dots, N\}$ pour $N \geqslant 1$ et $[x, y]_{\mathbb{Z}} = [x, y] \cap \mathbb{Z}$ pour $x, y \in \mathbb{R}$. Nous abrégeons parfois « progression arithmétique à k termes » par « k-PA ».

Lorsque $\mathbf{P}$ est une propriété, nous désignons par $1(\mathbf{P})$ le booléen qui vaut 1 lorsque $\mathbf{P}$ est vraie, et 0 sinon. Lorsque $\mathbf{P}_x$ est une propriété dépendant d'une variable x à valeurs dans un ensemble fini X, nous écrivons $\mathbb{P}_{x\in X}(\mathbf{P}_x) = \mathbb{E}_{x\in X} 1(\mathbf{P}_x)$.

Lorsque T est une quantité positive, nous utilisons la notation de Landau $O(T)$ (respectivement $\Omega(T)$) pour désigner une quelconque quantité inférieure à CT pour une constante $C > 0$ (respectivement une quantité supérieure[1] à cT pour une constante $c > 0$). Nous utilisons aussi la notation de Vinogradov : $U \ll V$ indique que $U = O(V)$, et $U \asymp V$ indique que l'on a simultanément $U \ll V$ et $V \ll U$.

[1] La notation Ω prend en général un sens légèrement différent en théorie analytique des nombres.

2. Bases : Combinatoire additive

Dans cette section, nous présentons brièvement le domaine de la combinatoire additive, avec un penchant assumé pour les aspects d'analyse harmonique du sujet. Un but secondaire est de mettre en place la notation utilisée à travers ce chapitre. Notre présentation est indéniablement influencée par les deux principaux ouvrages d'introduction à ce domaine [27, 100].

Quantités de combinatoire additive. Nous commençons notre exposition par l'un des points de départ de la combinatoire additive, qui est de réécrire des expressions combinatoires sous une forme analytique, qui peut être ensuite exploitée à l'aide de la transformée de Fourier. Pour cela, nous rappelons tout d'abord quelques notions rudimentaires d'analyse réelle [72]. Sauf mention du contraire, nous travaillons exclusivement avec un groupe abélien fini G, et ce pour le reste de ce chapitre ; nous supposons de plus que G n'a pas de 2-torsion par souci de simplicité.

Pour $p \geqslant 1$, nous définissons la norme L^p d'une function $f : G \to \mathbb{C}$ par

$$\|f\|_{L^p} = (\mathbb{E}_{x\in G}|f(x)|^p)^{1/p},$$

et nous écrivons $\|f\|_\infty = \sup_{x\in G} |f(x)|$. Puisque $\|f\|_{L^p} \to \|f\|_\infty$ lorsque $p \to \infty$, les normes L^p servent souvent à approcher les normes L^∞, qui peuvent être difficiles à estimer en pratique. Le produit scalaire de deux fonctions $f, g : G \to \mathbb{C}$ est défini par

$$\langle f, g\rangle = \mathbb{E}_{x\in G} f(x)\overline{g(x)}.$$

L'objet d'étude principal en combinatoire additive est un sous-ensemble fini de G, et nous souhaitons décrire celui-ci d'un point de vue fonctionnel. Étant donné un sous-ensemble A de G, on définit donc la fonction indicatrice 1_A en un point x

de G par

$$1_A(x) = 1(x \in A).$$

Remarquons que la densité de A s'obtient comme $|A|/|G| = \mathbb{E}_{x \in G} 1_A(x)$. Nous définissons aussi la fonction indicatrice normalisée de A par

$$\mu_A = \left(\tfrac{|A|}{|G|}\right)^{-1} \cdot 1_A,$$

de telle sorte que $\mathbb{E}\mu_A = 1$. On peut voir μ_A comme la densité de probabilité de la mesure de comptage sur A, que l'on écrit aussi μ_A : en effet l'on a $\mu_A(E) = \langle 1_E, \mu_A \rangle$ pour tout ensemble $E \subset G$.

DÉFINITION 2.1 (Convolution). *La convolution de deux fonctions $f, g : G \to \mathbb{C}$ est définie par*

$$f * g(x) = |G|^{-1} \sum_{u+v=x} f(u)g(v).$$

Cette opération est d'un intérêt immédiat pour l'étude de structures additives, comme nous l'expliquons à présent. Fixons trois sous-ensembles A, B, C de G pour les besoins de cette exposition. L'une des quantités les plus fondamentales en théorie des nombres additive est le nombre de représentations d'un élément $x \in G$ comme une somme $a + b$, où $(a, b) \in A \times B$. Après renormalisation, on obtient que cette quantité s'écrit

$$|G|^{-1} \#\{(a, b) \in A \times B : x = a + b\} = 1_A * 1_B(x).$$

Par conséquent, pour détecter la présence d'un ensemble structuré P dans un ensemble somme $A + B$, il suffit de vérifier que P est contenu dans le support de $1_A * 1_B$. Une autre configuration additive d'intérêt est la progression arithmétique à trois termes, définie ici comme un triplet $(x, x + d, x + 2d)$ où $x, d \in G$. Puisque nous avons supposé que G ne possède pas de 2-torsion, une progression (a, b, c) est

caractérisée par l'équation $a + c = 2b$, et le nombre normalisé de tels triplets dans $A \times B \times C$ est donné par

$$|G|^{-2}\#\{(a,b,c) \in A \times B \times C : a + c = 2b\} = \langle 1_A * 1_C, 1_{2\cdot B}\rangle, \tag{2.1}$$

où $2 \cdot B = \{2x,\, x \in B\}$. Cette simple expression se révèle très utile dans l'étude moderne du théorème de Roth. Une dernière quantité combinatoire importante est l'énergie additive de l'ensemble A, définie par

$$E(A) = \#\{(a_1,a_2,a_3,a_4) \in A^4 : a_1 + a_2 = a_3 + a_4\}.$$

En sommant sur $n = a_1 + a_2 = a_3 + a_4$, on voit que l'énergie normalisée de A est égale à

$$|G|^{-3}E(A) = \langle 1_A * 1_A, 1_A * 1_A\rangle. \tag{2.2}$$

Analyse harmonique discrète. Nous faisons à présent un bref survol de l'analyse de Fourier sur un groupe abélien fini, un outil qui se révèle d'une valeur capitale dans l'étude de certaines structures linéaires. La transformée de Fourier discrète est bien exposée dans [**100**, Section 4.1], et l'on peut en trouver une discussion plus approfondie dans [**27**, **45**]. Pour tous $x \in \mathbb{R}$ et $N \geqslant 1$, nous écrivons $e(x) = e^{2i\pi x}$ et $e_N(x) = e(x/N)$. Nous posons aussi $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ et $\mathbb{U} = \{z \in \mathbb{C} : |z| = 1\}$, de telle sorte qu'on a un isomorphisme canonique $\mathbb{T} \xrightarrow{\sim} \mathbb{U}$ donné par $\theta \mapsto e(\theta)$.

Introduisons comme précédemment un groupe abélien fini G, que l'on considère occasionnellement comme un $\mathbb{Z}$-module. Par le théorème de structure des groupes abéliens finis, on peut identifier G à un produit de groupes cycliques $\mathbb{Z}_{N_1} \times \cdots \times \mathbb{Z}_{N_d}$. Avec cette identification, on définit l'application $\cdot : G \times G \to \mathbb{T}$ par $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{d} \frac{x_i y_i}{N_i}$, et il est alors facile de vérifier que $\cdot$ est symétrique, $\mathbb{Z}$-bilinéaire et non dégénérée[2]. À l'aide de ces propriétés, il est aisé de déduire la propriété d'orthogonalité habituelle

[2] C'est-à-dire que pour tout $x \in G$, lorsque $x \cdot y = 0$ pour tout $y \in G$, on a $x = 0$.

des exponentielles, c'est-à-dire l'identité $\mathbb{E}_{x\in G}\, e(x\cdot y) = 1(y=0)$. La transformée de Fourier prend alors la forme suivante.

DÉFINITION 2.2 (Transformée de Fourier). *Soit $f : G \to \mathbb{C}$. Nous définissons*

$$\widehat{f}(r) = \mathbb{E}_{x\in G} f(x)e(-r\cdot x) \qquad\qquad (r \in G).$$

Lorsque $G = \mathbb{Z}_N$, nous avons l'expression explicite $\widehat{f}(r) = \mathbb{E}_{x\in\mathbb{Z}_N} f(x)e_N(-rx)$. Deux concepts fondamentaux de l'analyse harmonique sont l'inversion de Fourier, par laquelle on reconstitue la fonction d'origine à partir de ses coefficients de Fourier, et la formule de Parseval, une relation entre les produits scalaires sur l'espace physique et ceux sur l'espace des phases.

PROPOSITION 2.3 (Inversion de Fourier). *Soit $f : G \to \mathbb{C}$. On a*

$$f(x) = \sum_{r\in G} \widehat{f}(r)e(r\cdot x) \qquad\qquad (x \in G).$$

PROPOSITION 2.4 (Formule de Parseval). *Soient $f, g : G \to \mathbb{C}$. On a*

$$\langle f, g\rangle = \sum_{r\in G} \widehat{f}(r)\overline{\widehat{g}(r)}.$$

Remarquablement, la preuve des Propositions 2.3 et 2.4 est complètement élémentaire dans le cadre discret (puisqu'elle ne requiert que des échanges de sommation et la propriété d'orthogonalité des exponentielles), et ne présuppose aucune condition de régularité sur les fonctions utilisées. Cela constitue sans doute un attrait particulier de la combinatoire additive, et contraste fortement avec le cadre classique [57] où $G = \mathbb{T}$. Une dernière formule clé de l'analyse de Fourier est l'identité suivante, qui décrit le fait que convoluer dans l'espace physique revient à multiplier dans l'espace des phases.

PROPOSITION 2.5 (Identité de convolution). *Soient $f, g : G \to \mathbb{C}$. On a*

$$\widehat{f * g}(r) = \widehat{f}(r)\widehat{g}(r) \qquad\qquad (r \in G).$$

Une fonctionnalité importante de la convolution est son effet lissant sur les fonctions, qui est bien connue dans le cadre classique [57] : par exemple, la convoluée de deux fonctions de carré intégrable sur $\mathbb{T}$ est toujours continue. Il n'existe pas d'analogue définitif des notions de continuité ou de différentiabilité dans le cadre discret, cependant on peut raisonnablement interpréter comme une forme de « lissitude » le fait qu'une fonction f possède uniquement des petits coefficients de Fourier aux fréquences $r \neq 0$, par analogie avec le cas continu où une forte décroissance de la transformée de Fourier se traduit par une differentiabilité de grand ordre pour la fonction d'origine. Dans le cas de fonctions avec des coefficients de Fourier de module au plus 1, comme les fonctions indicatrices, la Proposition 2.5 montre alors que la convoluée de deux fonctions est plus lisse que celles d'origine.

Les Propositions 2.3–2.5 forment à elles trois le cœur de l'analyse de Fourier. Elles sont utilisées à répétition, et souvent implicitement, dans nos travaux des Chapitres III à V. Pour illustrer cette utilisation, nous obtenons ci-dessous des expressions harmoniques pour les quantités combinatoires vues précédemment. Pour commencer, le nombre normalisé de représentations d'un élément $x \in G$ comme une somme $a + b$, où $(a, b) \in A \times B$, prend la forme harmonique

$$1_A * 1_B(x) = \sum_r \widehat{1}_A(r)\widehat{1}_B(r)e(r \cdot x),$$

par inversion de Fourier et par l'identité de convolution. De même, par une application de la formule de Parseval et de l'identité de convolution à (2.1), nous pouvons réécrire le nombre normalisé de triplets en progression arithmétique dans $A \times B \times C$ comme

$$\langle 1_A * 1_C, 1_{2\cdot B}\rangle = \sum_{r\in G} \widehat{1}_A(r)\widehat{1}_B(-2r)\widehat{1}_C(r).$$

Finalement, l'énergie additive normalisée (2.2) devient

$$\langle 1_A * 1_A, 1_A * 1_A\rangle = \sum_{r\in G} |\widehat{1}_A(r)|^4.$$

Nous réinterprèterons cette identité plus tard dans la Section 4 comme décrivant l'égalité de la norme de Gowers U^2 d'un ensemble avec sa norme de Fourier ℓ^4.

3. Préliminaires : Analyse harmonique sur les ensembles de Bohr

Dans cette section nous introduisons les ensembles de Bohr, un outil technique majeur de la combinatoire additive, et nous expliquons les techniques modernes qui permettent de localiser l'analyse de Fourier à ces ensembles.

Ensembles de Bohr. Les ensembles de Bohr ont été popularisés par Ruzsa dans sa célèbre nouvelle preuve [77] du théorème de Freiman [17], un résultat pionnier de la combinatoire additive. Leur définition est donnée ci-dessous, où la notation $\|\cdot\| = d(\cdot, \mathbb{Z})$ désigne la pseudo-norme[3] habituelle sur le tore $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

DÉFINITION 3.1 (Ensemble de Bohr). *L'ensemble de Bohr ayant pour ensemble de fréquences $\Gamma \subset G$ et pour rayon $\delta > 0$ est*

$$B(\Gamma, \delta) = \{x \in G : \|r \cdot x\| \leqslant \delta \quad \forall r \in \Gamma\}.$$

Nous définissons la dimension de $B(\Gamma, \delta)$ par $d = |\Gamma|$.

Nous écrivons souvent de manière raccourcie B pour un ensemble de Bohr $B(R, \delta)$, et nous omettons parfois d'introduire les paramètres Γ, δ, d, qui sont alors supposés être implicitement définis. Si l'on considère l'application $(x, y) \mapsto x \cdot y$ comme un substitut de produit scalaire sur G (qui n'est pas un espace vectoriel), on peut voir un ensemble de Bohr comme un orthogonal approximatif de son ensemble de fréquences. Pour motiver cette définition, nous rappelons un concept important qui prend ses racines dans la théorie de dualité de Pontryagin [73].

DÉFINITION 3.2 (Annihilation). *Soient $\delta \in (0, 1]$ un paramètre, X un sous-ensemble de G et $R \subset G$ un ensemble de fréquences. On dit que R est η-annihilé*

[3] Par ce terme nous désignons le fait que $\|\cdot\|$ est définie-positive et satisfait l'inégalité triangulaire.

par X lorsque

$$|1 - e(r \cdot x)| \leqslant \eta \quad \textit{pour tous } x \in X \textit{ et } r \in R.$$

Ainsi, en se rappelant l'inégalité $|1 - e(y)| \leqslant 2\pi\|y\|$ valide pour tout $y \in \mathbb{T}$, on voit qu'un ensemble de Bohr $B(R, \delta)$ $2\pi\delta$-annihile toujours son ensemble de fréquences R. Cette dernière propriété est la plus importante en pratique, et pour cette raison elle est parfois prise comme la définition de ces ensembles. On peut se représenter visuellement un ensemble de Bohr comme un cube dans l'espace Euclidien : en effet un ensemble de Bohr d-dimensionnel n'est rien d'autre que le *pullback* de $[-\delta, \delta]^d$ par l'homomorphisme $x \mapsto (x \cdot r)_{r \in \Gamma}$. Nous définissons maintenant le ρ-dilaté d'un ensemble de Bohr $B = B(R, \delta)$ par $B_\rho = B(R, \rho\delta)$, ce que l'on peut comparer à l'opération de dilatation en géométrie Euclidienne. Une autre analogie est fournie par le comportement des ensembles de Bohr par rapport à l'addition : on a toujours $B_\rho + B_{\rho'} \subset B_{\rho+\rho'}$, comme pour la somme de Minkowski de cubes centrés dilatés. Finalement, les estimées standard [27] de croissance pour les ensembles de Bohr confirment à nouveau notre intuition géométrique.

PROPOSITION 3.3 (Taille et croissance des ensembles de Bohr). *Soit B un ensemble de Bohr d-dimensionnel de rayon $\delta \in (0, \frac{1}{2}]$. Alors*

$$|B| \geqslant \delta^d |G| \quad \textit{et} \quad |B_2| \leqslant 4^d |B|.$$

Régularité. Bien que les ensembles de Bohr aient l'avantage de se comporter comme un objet géométrique suffisamment simple, ils souffrent d'un défaut important : celui de n'être pas clos pour l'addition. Cela rend difficile, par exemple, l'estimation du nombre de progressions arithmétiques à trois termes dans un ensemble de Bohr. Pour contrer ce problème, Bourgain [5] a mis au point une preuve du théorème de Roth utilisant des ensembles de Bohr à différentes échelles, et a introduit la notion de régularité pour décrire les ensembles de Bohr qui sont pratiquement invariants par de petites dilations.

DÉFINITION 3.4 (Regularité). *On dit qu'un ensemble de Bohr d-dimensionnel B est régulier lorsque*

$$1 - 2^6\rho d \leqslant \frac{|B_{1\pm\rho}|}{|B|} \leqslant 1 + 2^6\rho d \qquad (0 < \rho < 2^{-6}/d).$$

Remarquons que lorsque B est un ensemble de Bohr d-dimensionnel et B' est un autre ensemble de Bohr contenu dans B_ρ, où $\rho \leqslant 2^{-6}/d$, on a

$$|(B + B') \triangle B| \ll \rho d|B|$$

(où $\triangle$ désigne la différence symmétrique), et l'on recouvre donc une forme de closure additive. Crucialement, un lemme de Bourgain [5] permet de travailler exclusivement avec des ensembles de Bohr réguliers en pratique.

PROPOSITION 3.5 (Régularisation des ensembles de Bohr). *Pour tout ensemble de Bohr B, il existe une constante $\kappa \in [\frac{1}{2}, 1]$ telle que B_κ soit régulier.*

Pour expliquer comment la régularité s'utilise en pratique, nous introduisons une nouvelle notation. Étant donné un paramètre $\varepsilon > 0$ et des quantités $X, Y \in \mathbb{C}$, on écrit $X \approx_\varepsilon Y$ pour indiquer que $|X - Y| \ll \varepsilon$. Nous fixons aussi un ensemble de Bohr d-dimensionnel B et un dilaté $\rho \leqslant 2^{-6}/d$. Nous calculons maintenant une quantité qui apparaît de manière récurrente dans les preuves du théorème de Roth : le nombre de progressions arithmétiques à trois termes dont les éléments appartiennent à des ensembles de Bohr à différentes échelles. La première étape est d'observer que pour tout $t \in B_\rho$, on a

$$\|\mu_{B+t} - \mu_B\|_{L^1} \leqslant \tfrac{|(B+t)\triangle B|}{|B|} \ll \rho d.$$

Lorsque λ est une mesure à support dans B_ρ, on a donc, par l'inégalité triangulaire,

$$\|\mu_B * \lambda - \mu_B\|_{L^1} = \|\mathbb{E}_{t\in G}\lambda(t)(\mu_{B+t} - \mu_B)\|_{L^1} \ll \rho d$$

Pour une telle mesure λ, et pour toute fonction f telle que $\|f\|_\infty \leqslant 1$, on a donc, par Hölder,

$$\langle f, \mu_B * \lambda\rangle \approx_{\rho d} \langle f, \mu_B\rangle.$$

Par suite, pour des ensembles $A \subset B$ et $A' \subset B_{\rho/2}$ arbitraires, on a

$$\langle 1_A * \mu_B, \mu_{2\cdot A'}\rangle = \langle 1_A, \mu_B * \mu_{2\cdot A'}\rangle \approx_{\rho d} \langle 1_A, \mu_B\rangle = |A|/|B|. \tag{3.1}$$

Après renormalisation, on en déduit qu'il y a environ $|A||A'|$ triplets en progression arithmétique dans $A \times A' \times B$. Le terme le plus à gauche de (3.1) est utile en pratique car il révèle l'opération de convolution sous-jacente, et car il peut-être directement transformé, par la formule Parseval, en l'expression harmonique

$$\langle 1_A * \mu_B, \mu_{2\cdot A'}\rangle = \langle \widehat{1}_A \cdot \widehat{\mu}_B, \widehat{\mu}_{2\cdot A'}\rangle.$$

Analyse spectrale locale. Le prochain sujet que nous abordons est l'analyse spectrale locale, une composante clé de deux de nos résultats résumés dans les Sections 6 et 7. Pour comprendre l'intérêt de cette analyse, considérons une fonction $f : G \to \mathbb{C}$, ainsi que sa série de Fourier $f(x) = \sum_r \widehat{f}(r)e(r \cdot x)$. En pratique, on peut souvent se permettre de tronquer cette somme, ainsi que d'autres expressions harmoniques plus compliquées, en négligeant la contribution des petits coefficients de Fourier. Il est alors critique d'analyser l'ensemble des fréquences restantes, que nous dénommons comme suit.

DÉFINITION 3.6 (Grand spectre). *Soient $\eta \in (0, 1]$ un paramètre et $f : G \to \mathbb{C}$ une fonction. Le η-spectre de f est*

$$\mathrm{Spec}_\eta(f) = \{r \in G : |\widehat{f}(r)| \geqslant \eta\|f\|_{L^1}\}.$$

Essayons maintenant de borner la taille du grand spectre d'un sous-ensemble X de G de densité τ. Par la borne de Tchebychev et la formule de Parseval, on obtient

$$|\operatorname{Spec}_\eta(1_X)| \leqslant (\tau\eta)^{-2} \sum_r |\widehat{1}_X(r)|^2 \leqslant (\tau\eta^2)^{-1}. \tag{3.2}$$

Nous affirmons qu'il est en général important d'obtenir un ensemble de Bohr qui annihile le grand spectre, et cela parce qu'un tel ensemble peut être utilisé dans l'étude du théorème de Roth ou des ensembles sommes, pour obtenir un incrément de densité ou pour construire un ensemble de presque-périodes comme expliqué dans les Sections 6 et 7. L'estimée (3.2) montre que si l'on choisit le spectre de X tout entier comme ensemble de fréquences, on peut l'annihiler par un ensemble de Bohr de dimension au plus $(\tau\eta^2)^{-1}$. Cependant, lorsque B est un ensemble de Bohr de dimension d et de rayon δ, on peut avoir $\tau \approx \delta^d$ et cette estimée est alors très faible. Une approche plus efficace a été conçue par Bourgain [**5**], qui a en fait prouvé un résultat structurel plus général, et la preuve du corollaire d'annihilation a été simplifiée par la suite par Green et Konyagin [**31**].

PROPOSITION 3.7 (Annihilation du spectre d'un ensemble de Bohr). *Soient $\varepsilon, \eta \in (0,1]$ des paramètres, et B un ensemble de Bohr d-dimensionnel régulier. Alors $\operatorname{Spec}_\eta(1_B)$ est ε-annihilé par B_ρ, à condition que $\rho \leqslant 2^{-7}\varepsilon\eta/d$.*

D'un autre côté, dans le contexte du théorème de Freiman-Ruzsa, il est souvent nécessaire d'annihiler efficacement le grand spectre d'un ensemble dense arbitraire. La célèbre borne de Chang [**7**] résout ce problème, et elle a trouvé rapidement une série d'applications aux problèmes de trouver des progressions arithmétiques dans les ensembles sommes [**9**, **29**] et des solutions d'équations linéaires non invariantes dans des ensembles denses [**87**, **89**], ainsi que dans des travaux subséquents sur le théorème de Freiman-Ruzsa [**88**].

PROPOSITION 3.8 (Borne de Chang). *Soient $\varepsilon, \eta \in (0,1]$ des paramètres. Soit X un sous-ensemble de G de densité τ. Le spectre $\mathrm{Spec}_\eta(1_X)$ est ε-annihilé par un ensemble de Bohr de dimension $d \ll \eta^{-2}\log\tau^{-1}$ et de rayon ε/d.*

La preuve de Chang est basée sur un ingénieux argument de dualité, qui repose lui-même crucialement sur une inégalité classique de Rudin [71]. Remarquons tout de même que la borne de Chang est assez inefficace comparée à celle de Bourgain lorsque X est un ensemble de Bohr d-dimensionnel de densité $b \approx \delta^d$, puisque dans ce cas la dimension de l'annihilateur obtenu est approximativement $\eta^{-2}d\log\delta^{-1}$, ce qui est bien supérieur à la dimension d'origine d. Sanders [78, 82] a par la suite développé un analogue local efficace de la borne de Chang, qui est devenu progressivement le nouveau standard dans les études du théorème de Roth [81, 90], de la théorie de Freiman-Ruzsa [83], et des progressions arithmétiques dans les ensembles sommes [51, 54].

PROPOSITION 3.9 (Annihilation du spectre local). *Soient $\varepsilon, \eta \in (0,1]$ des paramètres. Soit B un ensemble de Bohr régulier de dimension d et de rayon δ et X un sous-ensemble de B de densité τ. Alors $\mathrm{Spec}_\eta(1_X)$ est ε-annihilé par un ensemble de Bohr B' de dimension $d' \leqslant d+m$ et de rayon $\delta' \geqslant c\varepsilon\delta/d^2m$, où $m \ll \eta^{-2}\log\tau^{-1}$.*

4. Préliminaires : Analyse harmonique d'ordre supérieur

Dans cette section, nous rappelons quelques concepts de bases de la théorie de l'uniformité d'ordre supérieure, dont l'un des principaux objectifs est de mesurer jusqu'à quel point les sous-ensembles d'un groupe se comportent de manière pseudo-aléatoire, i.e. contiennent asymptotiquement le même nombre de configurations linéaires qu'un ensemble aléatoire de la même taille. Nous abordons aussi quelques aspects plus avancés de cette théorie qui concernent spécifiquement l'ensemble des nombres premiers.

Normes de Gowers. La nouvelle preuve analytique du théorème de Szemerédi [95] par Gowers [20] a introduit une classe de normes importante, qui

permet de mesurer en un certain sens les caractéristiques pseudo-aléatoires d'une fonction ; nous serons bientôt plus précis. À travers cette section, nous écrivons $\mathcal{C}$ pour l'opérateur de conjugaison sur $\mathbb{C}$, et $|\varepsilon| = \sum_i \varepsilon_i$ pour un vecteur $\varepsilon \in \{0,1\}^d$; nous abrégeons aussi $\mathbb{E}_{x\in G}$ par $\mathbb{E}_x$.

DÉFINITION 4.1 (Norme de Gowers). *Soit $f : G \to \mathbb{C}$ une fonction. Pour $d \geqslant 1$, la norme de Gowers U^d de f est*

$$\|f\|_{U^d}^{2^d} = \mathbb{E}_{x,u_1,\dots,u_d} \prod_{\varepsilon\in\{0,1\}^d} \mathcal{C}^{|\varepsilon|} f(x + \varepsilon_1 u_1 + \cdots + \varepsilon_d u_d). \tag{4.1}$$

La première de ces normes est d'une importance théorique moindre, mais il est d'usage de la définir pour initialiser certains arguments inductifs ; elle vaut

$$\|f\|_{U^1}^2 = \mathbb{E}_{x,u} f(x)\overline{f(x+u)} = |\mathbb{E}f|^2.$$

Remarquablement, la seconde de ces normes a une expression harmonique très simple :

$$\|f\|_{U^2}^4 = \mathbb{E}_{x,u,v} f(x)\overline{f(x+u)f(x+v)}f(x+u+v) = \langle f*f, f*f\rangle = \sum |\widehat{f}|^4$$

Malheureusement, pour $d \geqslant 3$, l'expression de Fourier de la norme U^d est beaucoup moins utile.

L'expression $\|f\|_{U^d}^{2^d}$ est une moyenne sur des parallélépipèdes discrets, et par conséquent elle satisfait plusieurs identités combinatoires remarquables. Pour décrire celles-ci, il est pratique d'introduire une nouvelle définition : la dérivée multiplicative d'une fonction f par rapport à un élément $u \in G$ est la fonction

$$\Delta_u f(x) = f(x+u)\overline{f(x)} \qquad (x \in G).$$

Remarquons que l'élevation au carré d'une moyenne a l'effet de dériver multiplicativement la fonction considérée :

$$|\mathbb{E}_x f(x)|^2 = \mathbb{E}_{x,y} f(x)\overline{f(y)} = \mathbb{E}_{u,y} f(y+u)\overline{f(y)} = \mathbb{E}_u \mathbb{E}_y (\Delta_u f)(y). \tag{4.2}$$

Dans le contexte du problème de Waring [104], cette technique très simple est connue sous le nom de dérivation de Weyl. Par induction, on peut aussi prouver la formule récursive suivante :

$$\|f\|_{U^{k+1}}^{2^{k+1}} = \mathbb{E}_u \|\Delta_u f\|_{U^k}^{2^k}. \tag{4.3}$$

Puisque $\|f\|_{U^1}^2 = |\mathbb{E} f|^2$, il s'ensuit par induction que le terme de droite de (4.1) est toujours positif. Par conséquent, la norme de Gowers $\|f\|_{U^d}$ est bien définie comme l'unique racine $1/2^d$-ème positive de cette expression. Montrer qu'il s'agit d'une vraie norme requiert plus de travail, et n'est en fait pas nécessaire pour la plupart des applications.

Ensembles pseudo-aléatoires. L'utilité des normes de Gowers dans l'étude du théorème de Szemerédi provient du fait qu'elles contrôlent, en un certain sens, les moyennes sur les progressions arithmétiques à un nombre fixé de termes : il s'agit là d'une observation clé de Gowers [20].

PROPOSITION 4.2 (Contrôle des k-PAs). *Soient $k \geqslant 2$ et des fonctions $f_1, \dots, f_k : G \to [-1,1]$. Alors, pour tout $1 \leqslant j \leqslant k$,*

$$|\mathbb{E}_{x,u} f_1(x) \cdots f_k(x+(k-1)u)| \leqslant \|f_j\|_{U^{k-1}}.$$

La preuve de cette proposition consiste en une série d'applications de l'inégalité de Cauchy-Schwarz, où chaque application élimine une fonction et dérive multiplicativement les autres, jusqu'à ce qu'il ne reste plus que la norme de Gowers de l'une d'entre elles. Essayons maintenant de comprendre comment cette proposition est appliquée en pratique pour estimer le nombre de progressions arithmétiques à k termes

dans un sous-ensemble A of G, sous l'hypothèse d'uniformité $\|1_A - \alpha\|_{U^{k-1}} = o(1)$, lorsque $|G| \to \infty$. Il est naturel d'introduire l'opérateur multilinéaire

$$T(f_1, \dots, f_k) = \mathbb{E}_{x,d} f_1(x) \cdots f_k(x + (k-1)d),$$

de telle sorte que le nombre de progressions arithmétiques à k termes dans A s'écrit $T(1_A, \dots, 1_A) \cdot |G|^2$. En écrivant $f_A = 1_A - \alpha$ pour la fonction balancée de A, et en développant $1_A = \alpha + f_A$ par multilinéarité, on obtient

$$T(1_A, \dots, 1_A) = \alpha^k + \sum T(*, \dots, f_A, \dots, *),$$

où la somme est sur $2^k - 1$ termes et les étoiles désignent des fonctions égales à α ou f_A. En appliquant la Proposition 4.2 à chaque terme de cette somme, nous pouvons en déduire l'énoncé suivant, où les termes $o(1)$ doivent être interprétés quand $|G| \to \infty$.

PROPOSITION 4.3 (Gowers-uniformité ⇒ comportement pseudo-aléatoire). *Soit A un sous-ensemble de G de densité α, et soit $f_A = 1_A - \alpha$. Si $\|f_A\|_{U^{k-1}} = o(1)$, alors A contient $(1 + o(1)) \cdot \alpha^k |G|^2$ arithmetic progressions à k termes.*

Remarquons que le nombre de progressions arithmétiques à k termes dans un sous-ensemble aléatoire A de G de densité α est asymptotiquement égal à $\alpha^k |G|^2$, puisque les évènements $x + id \in A$ où $0 \leqslant i < k$ sont alors approximativement indépendents et ont pour probabilité α. Par conséquent, les ensembles qui sont uniformes au sens de Gowers (c'est-à-dire, les ensembles A tels que $\|f_A\|_{U^{k-1}}$ est petit) se comportent de manière pseudo-aléatoire en termes du nombre d'occurences de k-PAs.

Complexité. Il s'avère que la Proposition 4.2 sur les moyennes sur les progressions arithmétiques de longueur donnée peut s'étendre à une classe bien plus large de configurations linéaires. Pour énoncer ces résultats, nous clarifions tout d'abord le vocabulaire que nous utilisons. Une forme linéaire entière est une application

$\varphi : \mathbb{Z}^d \to \mathbb{Z}$ de la forme $\varphi(x_1, \dots, x_d) = a_1x_1 + \cdots + a_dx_d$, où $a_1, \dots, a_d \in \mathbb{Z}$, qui peut être vue comme une forme linéaire sur $\mathbb{Q}^d$ pour tous les besoins d'algèbre linéaire. Un système de formes linéaires ψ est un uplet $(\psi_1, \dots, \psi_t)$, où $\psi_i : \mathbb{Z}^d \to \mathbb{Z}$ sont des formes linéaires ; nous supposerons toujours implicitement que les formes ψ_j sont distinctes. La notion de complexité de Cauchy-Schwarz (abrégée par CS-complexité dans la suite) introduite par Green et Tao [**39**] est alors la suivante.

DÉFINITION 4.4 (CS-Complexité). *Soit $\psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ un système de formes linéaires. La CS-complexité de ψ est le plus petit entier $s \geqslant 0$ tel que, pour tout $i \in [t]$, l'ensemble $\{\psi_j, j \neq i\}$ peut être partitionné en au plus $s+1$ classes disjointes, de façon à ce que ψ_i n'appartienne pas au sous-espace vectoriel engendré par l'une quelconque des classes. Lorsqu'un tel entier n'existe pas, on dit que le système a une complexité infinie.*

On peut vérifier que le système $\psi(x, d) = (x, x+d, \dots, x+(k-1)d)$ paramétrisant les progressions arithmétiques à k termes a une CS-complexité égale à $k-2$. La définition de CS-complexité n'est pas simple à manipuler, et pour contrôler les moyennes sur des configurations linéaires d'une CS-complexité donnée, il est préférable de mettre celles-ci sous une forme plus pratique, que l'on appelle la forme s-normale. Cela est expliqué convenablement dans la Section V.4, et nous nous contentons ici de dire qu'on peut toujours, en pratique, remplacer le système de formes d'origine par un système en forme normale. En développant la preuve de la Proposition 4.2, il est alors possible d'obtenir un contrôle des moyennes sur toute configuration linéaire de complexité finie à l'aide des normes de Gowers, comme le montre la proposition ci-dessous. Arrivé à ce point, nous spécialisons les énoncés à $G = \mathbb{Z}_M$ avec M un nombre premier ; en pratique M est choisi assez grand pour que la forme linéaire d'origine sur $\mathbb{Z}$ se réduise à une forme sur $\mathbb{Z}_M$ avec les mêmes propriétés de normalité.

PROPOSITION 4.5 (von Neumann généralisé, cas borné). *Soient $s \geqslant 0$ et $\psi : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ un système de formes linéaires sous forme s-normale. Soient $f_1, \dots, f_t : \mathbb{Z}_M^d \to [-1,1]$ des fonctions. Alors, pour tout $1 \leqslant j \leqslant t$,*

$$|\mathbb{E}_{n \in \mathbb{Z}_M^d} f_1(\psi_1(n)) \cdots f_t(\psi_t(n))| \leqslant \|f_j\|_{U^{s+1}}.$$

Cette proposition est un cas particulier d'un résultat de Green et Tao (comme expliqué dans [23]), et l'on peut en déduire facilement une estimée du nombre d'occurences d'une quelconque configuration linéaire de complexité finie dans un sous-ensemble A de $\mathbb{Z}_M$ suffisamment Gowers-uniforme, par le même argument que celui menant à la Proposition 4.3. Puisque nous avons $\|f\|_{U^2} = \|\widehat{f}\|_{\ell^4}$, les systèmes de CS-complexité 1 peuvent être analysés par des méthodes d'analyse harmonique classique, et cela constitue essentiellement le domaine d'applicabilité de ces méthodes. Une question profonde, posée par Gowers et Wolf [23], est de trouver la plus petite valeur de s pour laquelle la norme de Gowers U^{s+1} contrôle les moyennes de la forme $\mathbb{E}_{n \in \mathbb{Z}_M^d} f_1(\psi_1(n)) \cdots f_t(\psi_t(n))$, pour des fonctions f_i arbitraires bornées par 1 ; cette valeur est appelée la vraie complexité du système $(\psi_1, \dots, \psi_t)$. Pour notre travail du Chapitre V, qui traite principalement du cas des nombres premiers, nous n'avons pas besoin des résultats de la littérature florissante sur la vraie complexité [23–26, 37, 46, 47], mais nous soulignons qu'il s'agit là d'un sujet central de l'analyse harmonique d'ordre supérieur.

Moyennes linéaires sur les nombres premiers. Par contraste avec la situation précédente, pour détecter des configurations linéaires dans les nombres premiers, nous sommes forcés de travailler avec des fonctions non bornées, cousines de la fonction de von Mangoldt $\Lambda(n) = (\log n) 1(n = p^\nu)$. Pour recouvrer un théorème de type von Neumann pour de telles fonctions, Green et Tao [36, 39] ont d'abord construit un crible enveloppant. Il s'agit d'un poids $\nu : \mathbb{Z}_M \to \mathbb{R}^+$ majorant en chaque point les fonctions indicatrices des nombres premiers impliquées, et qui se comporte de manière pseudo-aléatoire au sens où, pour chaque système de formes

linéaires $\theta : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ de complexité finie,

$$\mathbb{E}_{n\in\mathbb{Z}_M^d}\nu(\theta_1(n))\cdots\nu(\theta_t(n)) = 1 + o(1)$$

lorsque $N \to \infty$, et où la vitesse de décroissance dépend de θ. (Il y a aussi une autre condition pour les systèmes θ contenant des formes linéaires identiques, que nous ne précisons pas.) La construction exacte de ν n'est pas importante pour notre discussion, et nous nous bornons à dire qu'elle est basée sur le même principe que le crible de Selberg. Green et Tao ont alors pu montrer que la Proposition 4.5 peut en effet être étendue aux fonctions qui sont simplement bornées par un poids pseudo-aléatoire, à l'aide d'un argument impliquant de nombreuses applications de Cauchy-Schwarz, et inspiré par des techniques de régularité sur les hypergraphes [**21**].

PROPOSITION 4.6 (von Neumann généralisé, cas pseudo-aléatoire). *Soient $s \geqslant 0$ et $\psi : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ un système de formes linéaires en forme s-normale. Soit $\nu : \mathbb{Z}_M \to \mathbb{R}^+$ un poids pseudo-aléatoire, et $f_1, \dots, f_t : \mathbb{Z}_M^d \to \mathbb{R}$ des fonctions telles que $|f_i| \leqslant \nu$ pour tout $i \in [t]$. Alors, pour tout $1 \leqslant j \leqslant t$, on a, lorsque $M \to \infty$,*

$$|\mathbb{E}_{n\in\mathbb{Z}_M^d} f_1(\psi_1(n))\cdots f_t(\psi_t(n))| \leqslant \|f_j\|_{U^{s+1}} + o(1).$$

Combiné avec un énoncé sur la décomposition de fonctions indicatrices des nombres premiers en une partie Gowers-uniforme et une partie se comportant comme un ensemble dense [**22**, Theorem 4.8] (et des considérations de réduction), ainsi qu'avec le théorème de Szemerédi, cette proposition est assez forte pour établir la présence de n'importe quelle configuration linéaire invariante par translation dans les premiers, et permet donc de reprouver le théorème de Green-Tao [**36**]. Il est bien plus difficile d'obtenir des asymptotiques pour ces configurations, ou plus généralement pour toutes les configurations affines de complexité finie, et cet objectif a finalement été atteint par Green, Tao et Ziegler [**39**–**42**] dans une série de papiers révolutionnaires dépassant 300 pages en volume. Pour notre travail résumé

à la Section 8, nous n'empruntons à nouveau que la Proposition 4.6 à cette vaste collection de travaux.

5. Préliminaires : Analyse harmonique sur les nombres premiers

Dans cette section, nous rappelons quelques résultats classiques de théorie des nombres multiplicative, puis nous nous intéressons à des estimées plus récentes sur les sommes exponentielles lacunaires de nombres premiers.

Notions de base. Nous commençons par rappeler les définitions standard du domaine [**103**]. On dit qu'une fonction $f : \mathbb{N} \to \mathbb{C}$ est multiplicative lorsque $f(1) = 1$ et $f(nm) = f(n)f(m)$ dès que $(n, m) = 1$. Un des intérêts de ces fonctions est que, par le théorème fondamental de l'arithmétique, elles sont déterminées par leur valeur aux puissances de nombres premiers : $f(\prod_i p_i^{\nu_i}) = \prod_i f(p^{\nu_i})$. Nous pouvons donc réécrire la somme d'une fonction multiplicative sur les entiers comme un produit Eulérien :

$$\sum_{n\geqslant 1} f(n) = \prod_p \sum_{\nu\geqslant 0} f(p^\nu)$$

à condition que, mettons, $\sum_p \sum_\nu |f(p^\nu)| < \infty$. L'extension de cette propriété à des fonctions multiplicatives de plusieurs variables est bien connue, et puisque de telles fonctions apparaissent dans notre travail, nous prenons le temps de décrire les formules correspondantes. Une fonction $F : \mathbb{N}^k \to \mathbb{C}$ est dite multiplicative lorsque $F(1, \dots, 1) = 1$ et

$$F(n_1 m_1, \dots, n_k m_k) = F(n_1, \dots, n_k)F(m_1, \dots, m_k)$$

dès que $(n_1 \dots n_k, m_1 \dots m_k) = 1$. Pour une telle fonction, on a

$$\sum_{n_1,\dots,n_k\geqslant 1} F(n_1, \dots, n_k) = \prod_p \sum_{\nu_1,\dots,\nu_k\geqslant 0} F(p^{\nu_1}, \dots, p^{\nu_k})$$

dès lors que le membre de gauche ou de droite est absolument convergent.

Rappelons aussi quelques résultats classiques sur la répartition des nombres premiers. Nous écrivons $\mathcal{P}$ pour l'ensemble de tous les nombres premiers, et $\pi(x) = |\mathcal{P} \cap [1, x]|$ pour la fonction de comptage des nombres premiers de Tchebychev. D'un point de vue combinatoire, la première chose à déterminer concernant l'ensemble $\mathcal{P}$ est sa densité dans un intervalle assez grand. La réponse à ce problème est fournie par le théorème des nombres premiers, prouvé indépendamment par Hadamard et de la Vallée-Poussin en 1896, et qui affirme que $\pi(x) \sim \frac{x}{\log x}$ lorsque $x \to \infty$. L'estimée $\pi(x) \asymp \frac{x}{\log x}$ était quand à elle déjà connue depuis 1851 grâce à Chebychev. Curieusement, cette estimée plus faible est la seule information sur la répartition des nombres premiers utilisée dans la preuve du célèbre théorème de Green et Tao [36] sur l'existence de progressions arithmétiques arbitrairement longues dans les nombres premiers.

Théorie du crible. Nous nous intéressons maintenant à un important domaine de la théorie analytique des nombres, celui de la théorie du crible. Le principe de toutes les méthodes de crible est de partir d'une suite arithmétique avec quelques propriétés de bonne répartition dans les classes de résidus, et d'en exclure tous les éléments satisfaisant un certain ensemble de conditions de divisibilité. On peut visualiser ce processus dans le cas du crible d'Eratosthène, où l'on considère les entiers de 1 à N, et l'on barre les entiers divisibles par 2, 3, 5, et ainsi de suite jusqu'à arriver au point où il ne reste plus que les nombres premiers dans $(\sqrt{N}, N]$ (ce n'est pas un crible très efficace en pratique [43]).

Pour donner les énoncés précis, nous devons tout d'abord rappeler le cadre formel d'un argument de crible [14]. Par une suite (finie) d'entiers, nous désignons ici un uplet $\mathcal{A} = (a_1, \dots, a_n) \in \mathbb{Z}^n$ où l'ordre est sans importance, et l'on écrit $\#\mathcal{A} = n$ pour le nombre d'éléments dans la suite. Étant donné un entier $d \geqslant 1$, nous écrivons aussi $\mathcal{A}_d = (a \in \mathcal{A} : d|a)$. Dans une situation de crible générique, on considère une suite $\mathcal{A}$ d'entiers, un ensemble $\mathfrak{P}$ de nombres premiers par lesquels cribler, et un seuil de criblage $z \geqslant 1$. Le nombre d'éléments non criblés de la suite

est alors

$$S(\mathcal{A}, \mathfrak{P}, z) = \#(a \in \mathcal{A} : p|a,\, p \in \mathfrak{P} \Rightarrow p > z).$$

On suppose de plus que, pour tout entier d sans facteurs carrés et à facteurs premiers dans $\mathfrak{P}$, on a

$$\#\mathcal{A}_d = \frac{\omega(d)}{d} X + r(d)$$

où $X \geqslant 1$, $\omega : \mathbb{N} \to \mathbb{R}^+$ est une fonction multiplicative et $r : \mathbb{N} \to \mathbb{R}$ doit être considéré comme un terme d'erreur. Le terme $\omega(p)$ représente intuitivement le nombre de classes modulo p que nous souhaitons exclure de la suite, et par conséquent nous supposons toujours que $0 \leqslant \omega(p) < p$ pour $p \in \mathfrak{P}$, afin de pouvoir trouver des survivants au processus de criblage. Puisque dans de nombreux travaux, $\omega(n)$ désigne le nombre de facteurs premiers de n, nous désignons ce dernier par $\nu(n)$ dans cette section. Une dernière quantité importante est le produit singulier

$$V(z) = \prod_{\substack{p \in \mathfrak{P} \\ p \leqslant z}} \Big(1 - \frac{\omega(p)}{p}\Big),$$

qui est exactement la probabilité « locale » qu'un résidu modulo $\prod_{p \in \mathfrak{P},\, p \leqslant z} p$ n'appartienne pas à $\omega(p)$ classes fixées modulo p, pour tout $p \in \mathfrak{P} \cap [2, z]$. Avec cette notation, nous pouvons maintenant énoncer un résultat central de la théorie du crible, sous une forme très simplifiée.

PROPOSITION 5.1 (Lemme fondamental de la théorie du crible). *Soient $\mathcal{A}$ et $\mathfrak{P}$ comme ci-dessus, et $v \geqslant 1$ et $\kappa > 0$ des paramètres. Supposons de plus que $\omega(p) \leqslant \kappa$ pour tout $p \in \mathfrak{P}$ et que $|r(d)| \leqslant \omega(d)$ pour tout entier d sans facteurs carrés et à facteurs premiers dans $\mathfrak{P}$. Alors, pour tout $z \geqslant 1$,*

$$S(\mathcal{A}, \mathfrak{P}, z) = \Big(1 + O(e^{-3v/2} v^{-v})\Big) \cdot XV(z) + O\bigg(\sum_{d < z^{2v}} 3^{\nu(d)} r(d)\bigg),$$

où la constante implicite dépend au plus de κ.

Cet énoncé exact est [14, Theorem 4.1], où il est déduit de l'élégant crible de Selberg. Le paramètre v est typiquement choisi assez petit pour que le terme d'erreur soit inférieur au terme principal. Déterminer la plus petite valeur de v pour laquelle cela est possible est une question centrale en théorie du crible, cependant pour de nombreuses applications on peut se permettre de choisir z comme étant une petite puissance de X, auquel cas la proposition ci-dessus suffit.

Transformée de Fourier des nombres premiers. Nous retournons maintenant à un point de vue de combinatoire additive, et nous nous demandons ce que l'on peut dire à propos de la transformée de Fourier d'un sous-ensemble des nombres premiers. Plus précisément, nous désignons par $\mathcal{P}_N$ l'ensemble des nombres premiers jusqu'à N, et nous considérons un sous-ensemble A de $\mathcal{P}_N$. Puisque $\mathcal{P}_N$ a une densité $\sim (\log N)^{-1}$ dans $[N]$ par le théorème des nombres premiers, il est naturel d'utiliser les fonctions normalisées

$$\lambda = (\log N) \cdot 1_{\mathcal{P}_N} \quad \text{et} \quad \lambda_A = (\log N) \cdot 1_A.$$

Nous sommes particulièrement intéressés par les moments $\|\widehat{\lambda}_A\|_p$ et $\|\widehat{\lambda}\|_p$ pour $p \geqslant 2$. Pour commencer, observons que par Plancherel et le théorème des nombres premiers, on a $\|\widehat{\lambda}\|_2 \asymp (\log N)^{1/2}$, et donc on ne peut espérer contrôler le second moment comme dans le cas des fonctions bornées. On peut tout de même obtenir la borne $\|\widehat{\lambda}\|_4 \ll 1$ via Plancherel et n'importe quel crible majorant (tel que celui de la Proposition 5.1). Par conséquent, le quatrième moment de $\widehat{\lambda}_A$ est lui aussi borné, grâce à Plancherel :

$$\|\widehat{\lambda}_A\|_4^4 = \langle \lambda_A * \lambda_A, \lambda_A * \lambda_A \rangle \leqslant \langle \lambda * \lambda, \lambda * \lambda \rangle = \|\widehat{\lambda}\|_4^4 \ll 1.$$

Cependant, dans le contexte du théorème de Roth, il est nécessaire de contrôler les moments $\|\widehat{\lambda}_A\|_p$ dans le domaine $p \in (2, 4)$, en partie parce que les moyennes

sur les progressions arithmétiques à trois termes sont bornées par

$$|\mathbb{E}_{x,d\in\mathbb{Z}_N} f_1(x)f_2(x+d)f_3(x+2d)| = |\sum_r \widehat{f_1}(r)\widehat{f_2}(-2r)\widehat{f_3}(r)| \leqslant \|\widehat{f_1}\|_3\|\widehat{f_2}\|_3\|\widehat{f_3}\|_3.$$

Notre travail résumé dans la Section 8 requiert aussi un contrôle satisfaisant d'un moment $\|\widehat{\lambda}_A\|_p$ avec $p \in (2,4)$.

Il s'avère que le problème d'estimer les moments $\|\widehat{\lambda}_A\|_p$ est lié à la propriété du majorant de Hardy-Littlewood en analyse harmonique : on dit qu'un sous-ensemble Λ de $[N]$ possède cette propriété pour $p > 0$ lorsque, pour toute suite $(a_n)_{n\in\Lambda}$ telle que $|a_n| \leqslant 1$, on a

$$\left\| \sum_{n\in\Lambda} a_n e(n\ \cdot) \right\|_{L^p(\mathbb{T})} \leqslant\ C(p) \left\| \sum_{n\in\Lambda} e(n\ \cdot) \right\|_{L^p(\mathbb{T})}.$$

Bourgain [3] a montré que l'ensemble des nombres premiers $\Lambda = \mathcal{P}_N$ satisfait cette propriété pour $p > 2$. En choisissant de plus $a_n = 1_A(n)$ ci-dessus, en renormalisant, et en utilisant un argument de discrétisation de Marcinkiewicz et Zygmund (voir [30, Lemma 6.5]), on peut déduire du résultat de Bourgain que pour tout $p > 2$,

$$\|\widehat{\lambda}_A\|_p \ll_p N^{1/p-1}(\log N)\left(\int_{\mathbb{T}} \left| \sum_{p\leqslant N} e(p\theta) \right|^p \mathrm{d}\theta \right)^{1/p}, \tag{5.1}$$

et nous sommes donc ramenés à étudier l'ensemble complet des nombres premiers. La somme exponentielle $\sum_{p\leqslant N} e(p\theta)$ et ses variantes avec poids sont un objet d'étude classique dans la méthode du cercle de Vinogradov [13, Chapter 25], par laquelle on peut montrer que le membre de droite de (5.1) est borné pour tout $p > 2$, et par conséquent $\|\widehat{\lambda}_A\|_p \ll_p 1$ pour tout $p > 2$.

Dans sa célèbre preuve du théorème de Roth dans les nombres premiers, Green [30] a obtenu une nouvelle preuve de la propriété du majorant pour les nombres premiers, à l'aide d'un argument inspiré par la théorie de la restriction, un domaine de recherche actif dont l'on peut trouver un très bon survol dans [58]. Green et Tao [34] ont ensuite découvert une approche plus générale et plus efficace pour ces estimées, en développant les arguments de restriction relativement à un crible

enveloppant développé par Ramaré [67] et par Ramaré et Ruzsa [68]. Pour être plus concret, nous présentons brièvement ce crible enveloppant ici, en suivant l'exposition de Green et Tao [34]. On considère un polynôme entier de la forme $F(X) = (a_1X + b_1)\cdots(a_kX + b_k)$, où $|a_i|, |b_i| \leqslant N$. On suppose que F n'a pas de diviseur premier fixe et que son discriminant est non nul, et l'on introduit un paramètre $C \leqslant R \leqslant N$. La définition exacte du crible enveloppant se révèle de peu d'importance pour les applications, mais nous la donnons ici pour la mettre en perspective : il s'agit de la fonction $\beta_R : \mathbb{N} \to \mathbb{R}^+$ définie en $n \geqslant 1$ par

$$\beta_R(n) = G(R)\Bigg(\sum_{\substack{d\leqslant R\\ d|F(n)}} \lambda_d^{\mathrm{SEL}}\Bigg)^2,$$

où λ_d^{SEL} sont les poids standard utilisés dans le crible de Selberg, et $G(R)$ est une certaine somme qui apparaît dans ce contexte (voir e.g. [14]). Crucialement, on peut montrer que

$$\beta_R(n) \gg \mathfrak{S}_F^{-1}(\log R)^k \cdot 1(\,p|F(n) \Rightarrow p > R\,) \qquad (n \in \mathbb{N}), \tag{5.2}$$

où

$$\mathfrak{S}_F = \prod_p \mathbb{P}_{n\in\mathbb{Z}_p}\big[p \nmid F(n)\big]\Big(1 - \frac{1}{p}\Big)^{-k}. \tag{5.3}$$

En particulier, cela signifie que β_R est non-nul aux entiers n tels que $F(n)$ est un nombre premier supérieur à R. L'intérêt principal du poids β_R réside dans la propriété du majorant suivante, implicite dans [34], et reprouvée régulièrement dans la littérature sur le principe de transférence.

PROPOSITION 5.2 (Propriété du majorant pour le crible enveloppant). *Soient $C \leqslant R \leqslant N^{1/10}$, et $f : [N] \to \mathbb{C}$ une fonction telle que $|f| \ll \beta_R$. Alors*

$$\|\widehat{f}\|_p \ll_p 1 \qquad (p > 2).$$

Cette estimée de restriction offre un grand degré de liberté dans les applications : grâce à celle-ci, Green et Tao [34] ont pu montrer que les nombres premiers de Chen contiennent une infinité de 3-PAs. Ce résultat a depuis été étendu par Tao [97] (dans un post de blog) à toute suite suffisamment dense d'entiers évitant un nombre fixés de classes modulo chaque petit nombre premier, en utilisant le grand crible au lieu du poids β_R de type Selberg.

Nous achevons cette section par une remarque importante, qui est que nous avons ignoré un aspect technique de taille : dans les applications au théorème de Roth, on doit en réalité considérer des sous-ensembles des nombres premiers dans une progression arithmétique modulo W, où $W = \prod_{p \leqslant \omega} p$ et ω croît lentement en fonction de N. En pratique, on peut toujours se ramener à cette situation par un argument basé sur le principe des tiroirs, et l'on adapte ensuite sans peine les fonctions indicatrices normalisées et les arguments de crible. Cela constitue le *W-trick* introduit par Green [30], qui a l'effet d'éliminer certaines obstructions de congruences locales, ce qui rend en particulier le poids β_R « pseudo-aléatoire sur les 3-PAs », au sens de la Section 4. Le *W-trick* joue un rôle important dans toutes les formes du principe du transfert [34, 50], et nous abordons cette question plus en détail à un stade ultérieur.

6. Résumé : Sur les progressions arithmétiques dans $A + B + C$

Dans cette section, nous exposons le principal résultat du Chapitre III, c'est-à-dire le Théorème III.1.6, en nous restreignant au cas où les trois ensembles sont identiques par simplicité. Nous cachons les facteurs logarithmiques en écrivant respectivement

$$\widetilde{O}\big(f(\alpha)\big) \quad \text{pour} \quad O\big(f(\alpha)(\log \alpha^{-1})^{O(1)}\big),$$
$$\widetilde{\Omega}\big(f(\alpha)\big) \quad \text{pour} \quad \Omega\big(f(\alpha)(\log \alpha^{-1})^{-O(1)}\big).$$

Nous écrivons aussi $\alpha \gtrsim f(N)$ pour une condition de la forme $\alpha \geqslant Cf(N)(\log N)^C$ où $C > 0$ est une constante non spécifiée.

Incrément de densité sur les ensembles de Bohr. Nous commençons par présenter la stratégie d'incrément de densité développée par Sanders [78] pour étudier les ensembles sommes, basée sur celle de Bourgain [5] pour le théorème de Roth, et qui forme le point de départ de la preuve de notre résultat principal. Supposons donc que A_0 est un sous-ensemble dense de $[M]$, et que nous souhaitons trouver une longue progression arithmétique dans $A_0 + A_0 + A_0$. La première chose à observer est que l'on peut plonger A_0 dans un groupe cyclique $\mathbb{Z}_N$ avec $N \sim 6M$ à l'aide de la projection $\pi : [M] \to \mathbb{Z}_N$, qui préserve les ensembles sommes triples et les progressions arithmétiques, et nous pouvons donc supposer que nous avons commencé avec un sous-ensemble A_0 de $\mathbb{Z}_N$ (quitte à réduire la densité d'origine par un facteur constant).

L'argument est inductif et à chaque étape on considère un ensemble de Bohr régulier (comme défini à la Section 3), ainsi qu'un sous-ensemble A de B de densité α. Par des techniques de régularité, il est toujours possible de trouver un ensemble de Bohr B' plus petit sur lequel un translaté A' de A a approximativement la même densité que A. Lorsque nous sommes dans le cas non structuré où $B' \not\subset A + A + A'$, l'intersection $\mathcal{U} = B' \cap (A + A + A')^c$ est non vide, et nous examinons le produit scalaire

$$\langle 1_A * 1_A * 1_{A'}, \mu_{\mathcal{U}} \rangle = 0.$$

En développant $1_A = \alpha 1_B + f_A$, et par certains calculs de régularité et en tronquant la série de Fourier, on a

$$\tfrac{1}{2}\alpha^2 b \leqslant \sum_{r:\, |\widehat{\mu}_{A'}(r)| \geqslant c\alpha} |\widehat{f}_A(r)|^2, \tag{6.1}$$

où b est la densité de B dans $\mathbb{Z}/N\mathbb{Z}$.

Par la stratégie d'incrément de densité ℓ^2, que nous décrivons dans la sous-section suivante, on peut alors obtenir un incrément de densité sur un autre ensemble de Bohr régulier, avec des nouveaux paramètres de densité, dimension et rayon

$$\alpha \leftarrow (1+c)\alpha, \quad d \leftarrow d + \widetilde{O}(\alpha^{-2}), \quad \delta \leftarrow \widetilde{\Omega}(1)\cdot\delta. \tag{6.2}$$

Puisque la densité ne peut excéder 1, l'itération se poursuit pendant au plus $\lesssim 1$ étapes, après lesquelles la dimension est $\lesssim \alpha^{-2}$ et le rayon est $\geqslant \exp[-\widetilde{O}(1)]$. Il est par ailleurs aisé de montrer qu'un ensemble de Bohr de dimension d et de rayon δ contient une progression arithmétique de longueur $\delta N^{1/d}$ [27], et de là on peut conclure que $A+A+A$ contient une progression arithmétique de longueur

$$\exp\left[\widetilde{\Omega}(\alpha^2)\log N\right] \quad \text{pourvu que} \quad \alpha \gtrsim (\log N)^{-1/2}.$$

Cette borne est de la qualité de notre théorème le plus « faible », c'est-à-dire le Théorème III.1.4, et donc pour abaisser la densité admissible nous devons faire appel à la machinerie développée par Sanders [81] pour obtenir des bornes très pointues dans le théorème de Roth.

Incrément de densité ℓ^2. La stratégie originale d'incrément de densité de Roth [69] exploite la grandeur d'un seul coefficient de Fourier, et procède par une itération sur des progressions arithmétiques. Szemerédi [96] et Heath-Brown [48] ont modifié cet argument pour exploiter à la place la grandeur d'un moment ℓ^2 de Fourier, et ils ont obtenu par là des bornes améliorées pour le théorème de Roth, en travaillant toujours avec des progressions arithmétiques. Bourgain [5] est l'inventeur de la stratégie d'incrément de densité ℓ^2 (et ℓ^∞) relative aux ensembles de Bohr, et l'analyse spectrale développée a posteriori par Sanders [78, 82], et exposée dans la Section 3, permet une généralisation importante de cette technique [81].

Afin de faciliter l'exposition, nous faisons un usage éhonté de notations très peu rigoureuses, mais très utiles. Nous désignons par $b, b', \ldots$ (respectivement $d, d', \ldots$) la densité (respectivement la dimension) d'ensembles de Bohr $B, B', \ldots$. Nous disons

qu'un ensemble $A \subset B$ a un incrément de densité de qualité $\alpha \leftarrow \alpha', d \leftarrow d', \delta \leftarrow \delta'$ lorsqu'il existe un autre ensemble de Bohr régulier B' de dimension d' et de rayon δ' sur lequel un translaté de A a pour densité α'. Finalement, nous écrivons $X \approx Y$ lorsque X et Y diffèrent d'une quantité « contrôlée » par un certain paramètre de régularité ρ, choisi assez petit en pratique ; c'est la notation la moins rigoureuse que nous employons. La stratégie d'incrément de densité ℓ^2 se résume alors essentiellement à l'énoncé suivant.

PROPOSITION 6.1 (Incrément de densité ℓ^2). *Soient $\nu, \eta \in (0,1]$ des paramètres. Soient B et $B' \subset B_\rho$ des ensembles de Bohr réguliers, A un sous-ensemble de B de densité relative α et X un sous-ensemble de B' de densité relative τ, où $\rho \leqslant c\nu\alpha/d$. Soit $f_A = 1_A - \alpha 1_B$, et supposons de plus que*

$$\sum_{r:\,|\widehat{\mu}_X(r)|\geqslant\eta} |\widehat{f}_A(r)|^2 \geqslant \nu\alpha^2 b. \tag{6.3}$$

Alors on a un incrément de densité de qualité

$$\alpha \leftarrow (1+c\nu)\cdot\alpha, \quad d \leftarrow d' + O(\eta^{-2}\log\tau^{-1}), \quad \delta \leftarrow (\eta/d')^2(\log\tau^{-1})^{-1}\cdot\delta.$$

Grâce à notre précédente présentation de la régularité et de l'analyse spectrale locale de la Section 3, il est maintenant aisé de prouver cetter proposition. En effet, observons tout d'abord que le domaine de sommation dans (6.3) est exactement égal à la quantité $\mathrm{Spec}_\eta(\mu_X)$ de la Définition 3.6. De la Proposition 3.9, nous déduisons que ce spectre est $\frac{1}{2}$-annihilé par un ensemble de Bohr $\dot{B}$ possédant la dimension et le rayon désiré, et l'on a $|\widehat{\mu}_{\dot{B}}(r)| = |\mathbb{E}_{x\in\dot{B}} e(r\cdot x)| \geqslant \frac{1}{2}$ pour tout $r \in \mathrm{Spec}_\eta(\mu_X)$. Mais alors

$$\nu\alpha^2 b \ll \sum_r |\widehat{f}_A(r)|^2 |\widehat{\mu}_{\dot{B}}(r)|^2 = \langle f_A * \mu_{\dot{B}}, f_A * \mu_{\dot{B}}\rangle \approx \|1_A * \mu_{\dot{B}}\|_2^2 - \alpha^2 b.$$

Par Hölder, nous avons donc

$$(1+c\nu)\cdot\alpha^2 b \ll \|1_A * \mu_{\dot{B}}\|_\infty \|1_A * \mu_{\dot{B}}\|_1 \ll \|1_A * \mu_{\dot{B}}\|_\infty \cdot \alpha b,$$

de telle sorte que A a une densité au moins égale à $(1 + c\nu) \cdot \alpha$ sur un translaté de $\dot{B}$, comme désiré. Dans la sous-section précédente, nous avons implicitement appliqué la Proposition 6.1 à (6.1) avec les paramètres $\nu \asymp 1$ et $\eta = c\alpha$.

La transformée de Katz-Koester généralisée et le lemme de Croot-Sisask. Afin d'obtenir la borne du Théorème III.1.6, nous devons combiner les deux principaux ingrédients du travail de Sanders [**81**] sur le théorème de Roth avec la stratégie d'incrément de densité améliorée utilisée dans la preuve du théorème $A + B$ dans son autre travail [**78**]. Le premier de ces ingrédients est le lemme de Croot-Sisask, un résultat d'une grande applicabilité prouvé dans [**11**], et que nous avons exposé en détail dans [**53**]. Le second est la transformée de Katz-Koester, qui a été en réalité développée par Sanders, qui attribue généreusement à Katz et Koester [**56**] une partie de l'inspiration derrière cet outil. Plus précisément, nous utilisons une généralisation de cet outil dûe à Bloom [**2**], et qui se révèle critique pour notre estimée de densité finale. Les énoncés précis sont donnés dans la Section III.5, et ici nous nous restreignons à expliquer en termes très informels ce que ces techniques apportent dans notre argument.

Nous commençons avec un sous-ensemble A de densité α d'un ensemble de Bohr régulier B. Comme précédemment, nous considérons un ensemble de Bohr B' à plus petite échelle et l'intersection A' d'un translaté de A avec B' ayant à peu près la même densité relative. Nous introduisons un nouveau paramètre $v \in (0, 1)$, et nous supposons que $A + A + A'$ a une densité inférieure à $1 - v$ dans B', de telle sorte que $\mathcal{U} = B' \cap (A + A + A')^c$ à une densité au moins égale à v dans B'. Comme auparavant, mais aussi avec quelques réarrangements de convolutions, on a

$$\langle 1_A * \mu_{A'} * \mu_{-\mathcal{U}}, 1_{-A} \rangle = 0.$$

Via la transformée de Katz-Koester généralisée, on peut transformer ce produit scalaire en

$$\langle 1_L * \mu_{S_1} * \mu_{S_2}, 1_{-A}\rangle = 0$$

où S_1, S_2 sont des sous-ensembles de densité $\exp[-\widetilde{O}(\alpha^{-1/2})\log v^{-1}]$ d'un ensemble de Bohr B'' plus petit, et L est un sous-ensemble de densité $\asymp 1$ de B. Par un lissage L^p de Croot-Sisask, on peut de plus obtenir un petit produit scalaire

$$\langle 1_L * \mu_{S_1} * \mu_{S_2} * \mu_X^{(\ell)} * \mu_{-X}^{(\ell)}, 1_{-A}\rangle \approx 0, \tag{6.4}$$

où X est un sous-ensemble de densité τ d'un ensemble de Bohr B''' encore plus petit et

$$\tau \geqslant \exp\Big[-\widetilde{O}(\alpha^{-1/2})\cdot \ell^2 \log v^{-1}\Big],$$

pour un paramètre $\ell \geqslant 1$. En développant $1_{-A} = f_{-A} + \alpha 1_B$ dans (6.4), et en tronquant la série de Fourier comme il est d'usage, puis par Cauchy-Schwarz, on peut obtenir l'inégalité

$$\alpha^2 b \ll \sum_r |\widehat{f}_A(r)|^2 |\widehat{\mu}_X(r)|^{4\ell}.$$

L'intérêt de l'opération de lissage de Croot-Sisask effectuée précédemment est que nous pouvons dorénavant nous restreindre à un spectre bien plus mince (et plus efficacement annihilé), avec $\ell \sim C\log\alpha^{-1}$:

$$\alpha^2 b \ll \sum_{r\in \mathrm{Spec}_{1/2}(\mu_X)} |\widehat{f}_A(r)|^2.$$

À ce stade (et à un autre que nous avons caché sous le tapis), la stratégie d'incrément de densité ℓ^2 intervient, et nous fournit un incrément de densité de qualité

$$\alpha \leftarrow (1+c)\cdot\alpha, \quad d \leftarrow d + \widetilde{O}(\alpha^{-1/2})\log v^{-1}, \quad \delta \leftarrow (v\alpha/d)^{O(1)}\cdot\delta.$$

Le gain en dimension par comparaison avec (6.2) est une conséquence de l'application de la transformée de Katz-Koester généralisée. Nous pouvons itérer ces bornes tant que B' a une densité inférieure à $1-v$ dans $A+A+A'$, et lorsque l'algorithme s'arrête nous avons donc trouvé, à l'intérieur d'un translaté de $A+A+A$, une proportion $1-v$ d'un ensemble de Bohr de dimension $d \lesssim \alpha^{-1/2}\log v^{-1}$ et de rayon $\delta \geqslant \exp[-\widetilde{\Omega}(1)\log v^{-1}]$. Finalement, un simple lemme combinatoire, aussi dû à Sanders [78], permet de trouver une PA de longueur v^{-1} dans cette portion d'un ensemble de Bohr à condition que $v \leqslant c\delta N^{1/d}/d$, et par une optimisation laborieuse du paramètre v, on peut faire en sorte que la PA soit de longueur

$$\exp\left[\widetilde{\Omega}(\alpha^{1/4})(\log N)^{1/2}\right] \quad \text{pourvu que} \quad \alpha \gtrsim (\log N)^{-2}.$$

7. Résumé : Progressions arithmétiques dans les ensembles à faible doublement

Dans cette section nous expliquons l'approche utilisée pour obtenir nos résultats du Chapitre IV.

Modélisation. Nous travaillons dans le cadre d'un groupe abélien quelconque G, et nous définissons une progression arithmétique à trois termes comme un triplet $(x, x+d, x+2d)$ avec $x, d \in G$, et nous disons que la progression est triviale lorsque $d = 0$. Étant donné un sous-ensemble fini A de G, nous cherchons à obtenir la plus grande valeur possible du paramètre de doublement $K = |A+A|/|A|$ pour laquelle on est sûr de pouvoir trouver une 3-PA non trivale dans A. Notre approche générale suit le principe bien connu de la modélisation [27], par lequel on réduit l'étude de A à celle d'un sous-ensemble dense d'un certain object structuré, auquel on peut ensuite appliquer une généralisation appropriée du théorème de Roth.

Une coset-progression est un ensemble de la forme

$$Q = [-N_1, N_1]_Z + \omega_1 + \cdots + [-N_d, N_d]_{\mathbb{Z}} \cdot \omega_d + H,$$

où $N_i \geqslant 1$, $\omega_i \in G$ et H est un sous-groupe de G. Cet objet apparaît naturellement dans l'énoncé par Green et Ruzsa [32] du théorème de Freiman-Ruzsa généralisé à un groupe abélien arbitraire. Un Freiman-s-isomorphisme entre deux sous-ensembles X et Y de groupes abéliens G et H est une application $\phi : X \to Y$ telle que, pour tous $(x_i), (x'_i) \in X^s$, on a

$$\sum_{i=1}^s x_i = \sum_{i=1}^s x'_i \quad \Leftrightarrow \quad \sum_{i=1}^s \phi(x_i) = \sum_{i=1}^s \phi(x'_i).$$

Pour $s \geqslant 2$, ces applications sont bijectives et préservent les progressions arithmétiques non triviales [27]. Une observation importante de [32] est que tout sous-ensemble fini d'un groupe abélien est Freiman-2-isomorphe à un sous-ensemble d'un groupe abélien *fini*, et donc pour le problème considéré nous pouvons travailler exclusivement dans ce type de groupes.

La technique de modélisation a été introduite par Ruzsa [77] dans le cadre des entiers, et se base sur le concept d'isomorphisme de Freiman [17] ; le lemme d'origine de Ruzsa a depuis été légèrement raffiné [9, 27]. Plus tard, Green et Ruzsa [32] ont obtenu un énoncé de modélisation plus général pour tout groupe abélien fini, qui est cependant coûteux dans les applications quantitatives, et dans notre situation nous avons besoin d'un résultat bien plus efficace de Sanders [83, Theorem 10.1].

PROPOSITION 7.1 (Modélisation de Sanders). *Soit A un sous-ensemble d'un groupe abélien fini tel que $|A + A| \leqslant K|A|$. Alors A a une densité au moins égale à $1/2K$ dans le translaté d'une coset-progression M régulière, d-dimensionnelle et telle que*

$$d \leqslant C(\log K)^6 \quad \text{et} \quad |M| \geqslant \exp\left[-C(\log K)^6(\log\log K)^6\right] \cdot |A|.$$

La présence du curieux adjectif « régulière » sera bientôt expliquée. La preuve de ce résultat est difficile, particulièrement sur le plan technique, et la présenter nous conduirait bien loin de notre objectif initial. Pour éviter cet écueil, nous référons plutôt le lecteur à la source d'origine [83], dont la compréhension peut

être grandement facilitée par la lecture du survol de Sanders [84]. Notre travail utilise la Proposition 7.1 essentiellement comme une boîte noire, et notre principale contribution technique est donc une extension des bornes de Sanders pour le théorème de Roth [81] aux systèmes de Bourgain, une catégorie d'ensembles qui inclut les coset-progressions. Un premier résultat de ce type avait déjà été obtenu par Sanders [80], et en exploitant la technologie de son travail sus-cité [81], nous avons pu obtenir l'estimée améliorée suivante.

Proposition 7.2 (Théorème de Sanders-Roth local). *Soient $\mathcal{B}$ un système de Bourgain d-dimensionnel régulier dans un groupe abélien fini sans 2-torsion*[4]*, et A un sous-ensemble de $\mathcal{B}$ de densité α. Alors le nombre de progressions arithmétiques à trois termes dans A est au moins*

$$\exp\left[-C(\alpha^{-1}+d)(\log d/\alpha)^7\right]\cdot|\mathcal{B}|^2.$$

Cela peut être combiné avec la Proposition 7.1 de la manière suivante. Soit A un ensemble de doublement $K \geqslant 1$ dans un groupe abélien fini sans 2-torsion, et soit M la coset-progression régulière donnée par la Proposition 7.1. Par la Proposition 7.2, le nombre de 3-PAs dans A est au moins

$$\exp\left[-CK(\log K)^7\right]\cdot|A|^2,$$

ce qui est supérieur au nombre $|A|$ de 3-PAs triviales dans A pour $K \leqslant (\log|A|)^{1-o(1)}$. Par les remarques précédentes, nous avons donc démontré que tout sous-ensemble fini A d'un groupe abélien (sans 2-torsion) de doublement au plus $(\log|A|)^{1-o(1)}$ contient une 3-PA non triviale, ce qui est pratiquement notre Théorème I.2 ! La difficulté principale réside donc dans l'obtention de la Proposition 7.2.

Systèmes de Bourgain. Une partie substantielle du Chapitre IV est dédiée à des rappels sur la notion de systèmes de Bourgain introduite par Green et

[4]Cette hypothèse peut être affaiblie, mais pas éliminée complètement.

Sanders [**33**], et à une description des analogues de la régularité et de l'analyse spectrale locale pour ces ensembles. Rappelons-nous que dans notre présentation de la Section 3, nous avons argumenté que les ensembles de Bohr se comportent de plusieurs façons comme des cubes dans l'espace Euclidien : la définition de système de Bourgain, qui est donnée précisément dans la Section IV.4, formalise ces propriétés géométriques.

Nous fixons maintenant un groupe abélien fini G, à l'intérieur duquel les ensembles que nous considérons par la suite sont contenus. Un système de Bourgain est une famille d'ensembles $\mathcal{B} = (B_\rho)_{\rho>0}$ qui satisfait certaines propriétés « de type cube », et où B_1 doit être considéré comme le cube principal (il est identifé à $\mathcal{B}$ dans la Proposition 7.2) et $(B_\rho)_\rho$ comme ses dilatés. Deux exemples importants de systèmes de Bourgain sont

$$\Big(B(\Gamma, \rho\delta)\Big)_{\rho>0} \qquad \text{(Système de Bohr)},$$

$$\Big(H \oplus_{i=1}^{d} [-\rho N_i, \rho N_i]_{\mathbb{Z}} \cdot \omega_i\Big)_{\rho>0} \qquad \text{(Système de coset-progression)}.$$

La preuve de la Proposition 7.2 consiste en une adaption aux systèmes de Bourgain de la stratégie d'incrément de densité de Sanders, que nous avons déjà rencontrée dans la Section 6. Puisque cette proposition est appliquée à la coset-progression de la Proposition 7.1, et puisque la stratégie d'incrément de densité remplace un système de Bourgain par son intersection avec un ensemble de Bohr annihilateur à chaque itération, les seuls systèmes de Bourgain que nous rencontrons dans notre argument sont des intersections de coset-progressions et d'ensembles de Bohr. S'il était possible de modéliser efficacement l'ensemble de départ de la Proposition 7.1 dans un ensemble de Bohr, il n'y aurait pas besoin de considérer de systèmes de Bourgain du tout, mais cela ne semble guère possible à partir de la preuve de ce résultat [**83**].

L'approche de Croot-Laba-Sisask. Notre travail dans le Chapitre IV contient aussi une extension d'un résultat de Croot, Laba et Sisask [9] sur l'existence de longues progressions arithmétiques dans tout sous-ensemble fini de $G = \mathbb{Z}$ à faible doublement, au cas d'un groupe abélien G arbitraire. Nous expliquons ici en termes informels notre version de la stratégie de Croot-Laba-Sisask, en insistant sur les endroits où nous avons apporté des modifications à leur argument.

Soit A un sous-ensemble d'un groupe fini abélien G, et supposons que $|A+A| \leqslant K|A|$ pour un paramètre de doublement $K \geqslant 1$. L'étape la plus importante dans l'approche de Croot-Laba-Sisask, et la seule que nous présentons en détail, est de trouver un ensemble de presque-périodes de la convolution $1_A * 1_A$. Dans ce contexte, u est une presque-période d'une fonction $f : G \to \mathbb{C}$ lorsque $\|\tau_u f - f\|_p \leqslant \frac{1}{2}\|f\|_p$; on pourrait exiger une différence moindre dans la définition, mais cela n'aurait que peu d'impact sur notre argument.

Par la version de Sanders [83] du lemme de Bogolyubov-Ruzsa (une variante de la Proposition 7.1, dont on peut la déduire), on peut trouver un grand système de Bourgain $B \subset 2A - 2A$, et d'après l'inégalité de Petridis-Plünnecke-Ruzsa [27], on a $|A+B| \leqslant K^5|A|$. En exploitant cette structure additive à l'aide du lemme de Croot-Sisask, on peut approximer la convolution $1_A * 1_A$ en norme L^p par $1_A * 1_A * \lambda_X^{(\ell)}$, où $\ell \geqslant 1$ est un paramètre, $\lambda_X = \mu_X * \mu_{-X}$ et X est un sous-ensemble relativement dense de B.

Nous pouvons maintenant choisir, pour l'ensemble de presque-périodes recherché, n'importe quel système de Bourgain $\widetilde{\mathcal{B}}$ qui ε-annihile le demi-spectre de X, et en particulier celui de l'analogue approprié de la Proposition 3.9. En effet, lorsque u appartient à un tel ensemble $\widetilde{\mathcal{B}}$ on a, par la séparation de sommes usuelle par

rapport au petit/grand spectre,

$$\|1_A * 1_A * \lambda_X^{(\ell)} - \tau_u 1_A * 1_A * \lambda_X^{(\ell)}\|_\infty \leqslant \sum_r |\widehat{1}_A(r)|^2 |\widehat{\mu}_X(r)|^{2\ell} |1 - e(r \cdot u)|$$
$$\ll \varepsilon \sum_{|\widehat{\mu}_X(r)| \geqslant 1/2} |\widehat{1}_A(r)|^2 + 2^{-2\ell} \sum_{|\widehat{\mu}_X(r)| \leqslant 1/2} |\widehat{1}_A(r)|^2,$$

qui peut être rendu aussi petit que nécessaire en choisissant ℓ grand et ε petit. Puisque la norme ℓ^∞ contrôle la norme ℓ^p pour les fonctions à support étroit, et puisque $1_A * 1_A * \lambda_X^{(\ell)}$ est proche de $1_A * 1_A$ en norme L^p, on peut finalement conclure que $\widetilde{\mathcal{B}}$ est un ensemble de presque-périodes de cette dernière fonction.

Par un lemme de concentration de Croot, Laba et Sisask [9], on peut déduire de ce qui précède que tout sous-ensemble « pas trop grand » de l'ensemble des presque-périodes $\widetilde{\mathcal{B}}$ est contenu à translation près dans le support de $1_A * 1_A$, c'est-à-dire dans $A + A$, et par un simple argument de dilatation on peut choisir ce sous-ensemble comme étant une progression arithmétique ou un sous-groupe de taille raisonnable. Cela implique d'optimiser les différents paramètres entrant en jeu, ce qui est moins intéressant d'un point de vue conceptuel, et nous ne discutons donc pas cette partie de l'argument plus en détail.

Par comparaison, l'argument d'origine de Croot, Laba et Sisask [9] concernait un ensemble d'entiers A de doublement K, et dans ce cas on peut supposer que A est contenu dans un groupe cyclique $\mathbb{Z}_N$ où il a une densité K^{-C} et le même doublement, à l'aide du lemme de modélisation de Ruzsa [27]. Au lieu du lemme de Bogolyubov-Ruzsa, on utilise alors l'estimée de doublement par densité $|A + \mathbb{Z}_N| \leqslant K^C |A|$, et puisque l'ensemble X à annihiler vit dans $\mathbb{Z}_N$, une application de la borne de Chang (Proposition 3.8) suffit. Bien que très proche conceptuellement, l'argument analogue pour les systèmes de Bourgain apporte de légères complications techniques.

8. Résumé : Sur les systèmes de complexité un dans les nombres premiers

Dans cette section, nous exposons notre travail le plus récent, qui constitue le Chapitre V de cette thèse. Nous expliquons d'abord la structure de notre preuve, puis nous discutons en détail deux arguments empruntés à la littérature que nous utilisons dans notre argument.

Principe du transfert. Soit $\psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ un système de formes linéaires de complexité un au sens de la Section 4, avec la condition supplémentaire d'invariance par translation : $(u_1, \dots, u_t) \in \mathrm{Im}(\psi) \Rightarrow (u_1 + h, \dots, u_t + h) \in \mathrm{Im}(\psi)$ pour tout $h \in \mathbb{Z}$. Notre objectif est d'identifier une configuration $\psi(x) \in A_0^t$ à coordonnées distinctes, lorsque A_0 est un sous-ensemble des nombres premiers jusqu'à N de densité $\alpha \geqslant C(\log\log N)^{-c}$, pour un certain $c = c(\psi)$. Notre stratégie principale suit le principe du transfert familier aux experts, par lequel on passe d'un sous-ensemble dense des nombres premiers à un sous-ensemble dense des entiers, auquel on peut appliquer un théorème de type Szemerédi pour trouver la configuration désirée.

Nous présentons maintenant cet argument de manière plus précise, et nous commençons par appliquer le *W-trick*, une réduction standard qui nous permet de remplacer l'ensemble d'origine A_0 par un sous-ensemble A de $[N]$ de densité $\alpha\xi(W)(\log N)^{-1}$ tel que $b + W \cdot A \subset \mathcal{P}$, où $W = \prod_{p \leqslant \omega} p$, b est un entier premier à W et $\xi(n) = n/\phi(n)$. Puisque notre argument repose en partie sur le travail de Helfgott et de Roton [50], nous devons choisir un large module $\omega \sim c \log N$.

En conséquence, nous utilisons une fonction indicatrice normalisée

$$\lambda_A = \xi(W)^{-1}(\log N) \cdot 1_A$$

de telle sorte que $\mathbb{E}\lambda_A = \alpha$. Nous introduisons aussi une nouvelle échelle $M \sim CN$ et nous considérons les fonctions sur $\mathbb{Z}$ telles que λ_A comme des fonctions sur

$\mathbb{Z}_M$ (en un sens approprié). Enfin, nous introduisons l'opérateur de comptage de configurations défini par

$$T(f_1, \dots, f_t) = \mathbb{E}_{n \in \mathbb{Z}_M^d} f\Big[\psi_1(n)\Big] \dots f_t\Big[\psi_t(n)\Big].$$

Notre point de départ est le principe du transfert de Helfgott et de Roton [**50**], qui améliore celui de Green [**30**], et qui était à l'origine conçu pour le cas des 3-PAs. Le transfert en question consiste à comparer le compte $T(\lambda_A, \dots, \lambda_A)$ au compte $T(\lambda'_A, \dots, \lambda'_A)$, où λ'_A est une approximation de λ_A en norme U^2 qui se comporte essentiellement comme un sous-ensemble de $\mathbb{Z}_M$ de densité α^2. Plus précisément, le machinerie de Helfgott-de Roton fournit la borne

$$\|\lambda_A - \lambda'_A\|_{U^2} \ll (\log N)^{-c} \tag{8.1}$$

dans le domaine $\alpha \geqslant C(\log\log N)^{-c}$, et montre que l'ensemble niveau $\{\lambda'_A \geqslant \alpha/2\}$ a une densité au moins égale à $c\alpha^2$ dans $\mathbb{Z}_M$.

Nous pouvons ensuite développer $\lambda_A = \lambda'_A + (\lambda_A - \lambda'_A)$ par multilinéarité pour obtenir

$$T(\lambda_A, \dots, \lambda_A) = T(\lambda'_A, \dots, \lambda'_A) + \sum T(*, \dots, \lambda_A - \lambda'_A, \dots, *), \tag{8.2}$$

où les étoiles désignent des fonctions égales à λ'_A où $\lambda_A - \lambda'_A$, et la somme doit être interprétée comme un ensemble de termes d'erreurs. En appliquant notre extension (Proposition V.8.1) du théorème de type Szemerédi de Shao à la fonction-ensemble λ'_A, nous pouvons estimer le terme principal par $T(\lambda'_A, \dots, \lambda'_A) \geqslant \exp[-C\alpha^{-C}]$. En supposant pour l'instant que les termes d'erreur dans (8.2) sont $\ll (\log N)^{-c}$, nous pouvons conclure que $T(\lambda_A, \dots, \lambda_A) \geqslant \exp[-C\alpha^{-C}]$ dès que $\alpha \geqslant C(\log\log N)^{-c}$. Puisque $\lambda_A \leqslant (\log N) \cdot 1_A$, cela nous dit que nous pouvons trouver une large quantité de configurations $\psi(x) \in A^t$, et en particulier une qui n'est pas triviale.

Les termes d'erreur sont estimés à l'aide de notre version quantifiée (Proposition V.6.4) du théorème de von Neumann généralisé de Green et Tao [**39**]. Cette

version dit que lorsque des fonctions $f_1, \dots, f_t : \mathbb{Z}_M \to \mathbb{R}$ sont bornées en chaque point par un poids $\nu : \mathbb{Z}_M \to \mathbb{R}^+$ pseudo-aléatoire de niveau H, on a

$$|T(f_1, \dots, f_t)| \leqslant \|f_i\|_{U^2} + O(H^{-1/4}) \qquad (1 \leqslant i \leqslant t). \tag{8.3}$$

Nous disons ici qu'un poids $\nu : \mathbb{Z}_M \to \mathbb{R}^+$ est pseudo-aléatoire de niveau H lorsque sa moyenne sur toute configuration linéaire $\theta : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ de complexité finie est égale à $1 + O(H^{-1})$ (essentiellement). Les fonctions λ_A et λ'_A sont majorées par une version moyennée du poids de GPY

$$\Lambda_{\chi,R,W}(n) = \xi(W)^{-1}(\log R) \cdot \Big(\sum_{d | Wn+b} \mu(d)\gamma(d) \Big)^2,$$

où $\gamma(d) \geqslant 0$ sont des réels bien choisis et à support sur $d \leqslant R$.

Une partie substantielle de notre argument consiste alors à prouver que $\Lambda_{\chi,R,W}$ est pseudo-aléatoire de niveau $H = (\log N)^c$ sous le régime $\omega \sim c \log N$, qui nous est imposé par l'utilisation de la technologie de Helfgott et de Roton. En substituant cette valeur de H et (8.1) dans (8.3), nous pouvons borner les termes d'erreurs de (8.2) par la quantité désirée.

Principe du transfert de Helfgott et de Roton. Notre argument utilise de façon critique le principe de transfert de Helfgott et de Roton [**50**], et il nous semble donc approprié de donner un bref survol de cette technique.

Nous commençons par décrire la construction d'une approximation λ'_A de λ_A en norme U^2. Pour le poids β_R de la Section 5 et les choix $F(X) = WX + b$ et $R = N^{1/20}$, et en se rappelant la définition (5.3), on a

$$\mathfrak{S}_F = \prod_p \mathbb{P}_{n \in \mathbb{Z}_p}\big[p \nmid Wn + b\big]\Big(1 - \frac{1}{p}\Big)^{-1} \asymp \xi(W).$$

Puisque $\lambda_A = \xi(W)^{-1}(\log N) \cdot 1_A$, on déduit de (5.2) avec $k = 1$ que $0 \leqslant \lambda_A \ll \beta_R$ en tout point. Par la propriété du majorant de la Proposition 5.2, il s'ensuit que $\|\widehat{\lambda}_A\|_q \ll_q 1$ pour tout $q > 2$. Nous définissons maintenant $\lambda'_A = \lambda_A * \mu_B$, où B est un

ensemble de Bohr annihilant le grand spectre de λ_A (nous ignorons les paramètres précis entrant en jeu). Par l'expression de Fourier de la norme U^2, nous avons donc

$$\|\lambda_A - \lambda'_A\|_{U^2}^4 = \sum_r |\widehat{\lambda}_A(r)|^4 |1 - \widehat{\mu}_B(r)|^4.$$

En séparant comme d'habitude les sommes sur le petit/grand spectre, on peut rendre cette quantité aussi petite que nécessaire, en utilisant de manière cruciale le fait que l'un des moments $\|\widehat{\lambda}_A\|_p$ avec $p \in (2,4)$ est borné.

Comme nous l'avons déjà vu, la prochaine étape de la stratégie de transfert de Helfgott et de Roton [50] est de montrer que l'ensemble niveau $\{\lambda'_A \geqslant \alpha/2\}$ a une densité $\gg \alpha^2$ dans $\mathbb{Z}_M$. Cela s'obtient en considérant le second moment

$$\|\lambda_A * \mu_B\|_2^2 = \mathbb{E}_{m_1,m_2 \in B} \lambda_A(n+m_1)\lambda_A(n+m_2),$$

et l'on peut montrer que ce moment est borné à l'aide d'un crible majorant, à condition que B soit assez grand : cela impose la restriction précédente $\alpha \geqslant (\log\log N)^{-c}$. Nous n'expliquons pas cette étape plus en détail, si ce n'est pour dire qu'il s'agit là du moment précis où l'on a besoin de fixer $\omega \sim c \log N$. Finalement, un lemme de concentration de Helfgott et de Roton [50] énonce que lorsqu'une fonction $f : \mathbb{Z}_M \to \mathbb{R}^+$ est telle que $\mathbb{E}f = \alpha$ et $\|f\|_2 \ll 1$, l'ensemble niveau $\{f \geqslant \alpha/2\}$ a une densité au moins égale à $c\alpha^2$ dans $\mathbb{Z}_M$, et nous pouvons appliquer cela à $f = \lambda'_A$. Le travail de Naslund [65] optimise ces deux dernières étapes, et permet d'obtenir l'exposant $1 + o(1)$ au lieu de 2 dans la densité de l'ensemble niveau ; dans notre travail cela a l'effet de préserver l'exposant des théorèmes de type Szemerédi dans les entiers au cas des nombres premiers, modulo la perte habituelle d'un logarithme.

Le théorème local inverse U^2 de Shao. Dans notre extension modeste du théorème de type Szemerédi de Shao [91], du cas des d-configurations à celui des configurations linéaires arbitraires de complexité un, nous invoquons le théorème local inverse U^2 prouvé dans ce travail. Par souci de complétude, nous esquissons

aussi les idées principales derrière ce résultat, en essayant de le voir sous un jour plus Fourier-analytique (bien que la preuve soit, en substance, exactement la même).

Nous définissons tout d'abord, pour $g : \mathbb{Z}_M \to \mathbb{R}$ et des sous-ensembles $X_1, X_2 \subset \mathbb{Z}_M$,

$$\|g\|^4_{\boxtimes(X_1\times X_2)} = \mathbb{E}_{x_1,x_1'\in X_1}\mathbb{E}_{x_2,x_2'\in X_2}\, g(x_1+x_2)g(x_1+x_2')g(x_1'+x_2)g(x_1'+x_2').$$

La norme locale U^2 d'une fonction $f : \mathbb{Z}_M \to \mathbb{R}$ relativement aux ensembles $X_0, X_1, X_2 \subset \mathbb{Z}_M$, telle que définie par Shao [91], est alors

$$\|f\|^4_{U^2(X_0,X_1,X_2)} = \mathbb{E}_{x\in X_0}\|f(x_0+\cdot)\|^4_{\boxtimes(X_1\times X_2)}.$$

Le problème est dorénavant le suivant : étant donné trois ensembles de Bohr réguliers B, B', B'' tels que $B \leqslant_\rho B'$ et $B' \leqslant_\rho B''$ pour un petit dilaté ρ, et une fonction $f : \mathbb{Z}_M \to \mathbb{C}$ telle que $\mathbb{E}_B f = 0$, que peut-on dire de f lorsque $\|f\|_{U^2(B,B',B'')} \geqslant \eta$, pour un certain paramètre $\eta \in (0,1]$? Dans notre situation, nous souhaitons de fait montrer que, lorsque $A \subset B$ et $f = 1_A - \alpha 1_B$, on peut obtenir un incrément de densité sur un ensemble de Bohr plus petit.

La première étape est de trouver, par regularité et par le principe des tiroirs, un élément $x_0 \in B_{1-\rho}$ tel que $|\mathbb{E}_{B'} f|$ est petit et

$$\eta^4 \ll \|f(x_0+\cdot)\|^4_{\boxtimes(B'\times B'')}.$$

En écrivant $g(x) = f(x_0+x)1_{B'+B''}(x)$, on a, par régularité et après renormalisation,

$$\eta^4 b'^{\,2} \ll \mathbb{E}_{x_1,x_1'\in\mathbb{Z}_M}\mathbb{E}_{x_2,x_2'\in\mathbb{Z}_M} g(x_1+x_2)g(x_1+x_2')g(x_1'+x_2)g(x_1'+x_2')\mu_{B''}(x_2)\mu_{B''}(x_2').$$

Par la transformée de Fourier, cela devient

$$\eta^4 b'^{\,2} \ll \sum_{r,s} |\widehat{g}(r)|^2|\widehat{g}(s)|^2|\mu_{B''}(r+s)|^2.$$

La somme sur r peut être bornée en ℓ^∞–ℓ^1, et la somme de Fourier tronquée, après quoi on obtient

$$\eta^4 b' \ll \max_r \sum_{-r+\mathrm{Spec}_{\eta^2}(\mu_{B''})} |\widehat{g}(s)|^2.$$

Choisissons maintenant r atteignant ce maximum. Pour annihiler le domaine de sommation, il est alors suffisant d'annihiler simultanément la fréquence $-r$ et le grand spectre de B'', ce qui d'après les estimées de la Section 3 est possible simplement en dilatant B'' et en ajoutant $-r$ à son ensemble de fréquences. Lorsque $A \subset B$ et $f = 1_A - \alpha 1_B$, on peut finalement obtenir, à l'aide de la stratégie d'incrément de densité ℓ^2 et de plusieurs calculs de régularité additionnels, un incrément de densité de qualité

$$\alpha \leftarrow (1 + c\eta^8)\alpha, \quad d \leftarrow d + 1, \quad \delta \leftarrow (\eta\rho/d)^{O(1)} \cdot \delta.$$

Chapitre III. On arithmetic progressions in $A+B+C$

Author: Kevin Henriot.

Abstract: Our main result states that when A, B, C are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities α, β, γ, the sumset $A+B+C$ contains an arithmetic progression of length at least $e^{c(\log N)^c}$ for densities $\alpha \geqslant (\log N)^{-2+\varepsilon}$ and $\beta, \gamma \geqslant e^{-c(\log N)^c}$, where c depends on ε. Previous results of this type required one set to have density at least $(\log N)^{-1+o(1)}$. Our argument relies on the method of Croot, Laba and Sisask to establish a similar estimate for the sumset $A+B$ and on the recent advances on Roth's theorem by Sanders. We also obtain new estimates for the analogous problem in the primes studied by Cui, Li and Xue.

1. Introduction

Let A and B be subsets of a cyclic group $\mathbb{Z}/N\mathbb{Z}$ of density α and β. The problem of finding long arithmetic progressions in $A+B$ has a rich history starting with the striking result of Bourgain [4]: the sumset $A+B$ always contains an arithmetic progression of length at least $e^{c(\alpha\beta\log N)^{1/3}}$ provided the densities satisfy $\alpha\beta \geqslant (\log N)^{-1+o(1)}$ (and the progression is nontrivial in this range: this will always be the case later when we specify a range of density). Major progress was made by Green [29] who showed that, under the same condition on densities, the progression could be taken as large as $e^{c(\alpha\beta\log N)^{1/2}}$. Sanders [78] later found a very different proof of Green's theorem and yet a third and relatively simple proof was provided recently by Croot, Laba and Sisask [9].

For fixed densities α and β, the progression found has length $e^{c\sqrt{\log N}}$ and this has not been improved to date, while a negative result of Ruzsa [74] says that one cannot do better than $e^{c(\log N)^{2/3+\varepsilon}}$. However when densities are allowed to decrease with N, a remarkable result was obtained recently by Croot, Laba and Sisask [9]. Improving on a first result of Croot and Sisask [11], they showed that the sumset $A+B$ contains an arithmetic progression of size at least $e^{c(\alpha\log N)^{1/2}/(\log 2\beta^{-1})^{3/2}}$ in a range $\alpha(\log\frac{2}{\beta})^{-5} \geqslant C(\log N)^{-1+o(1)}$. While the theorems of Bourgain and Green require one set to have density at least $(\log N)^{-1/2+o(1)}$, this allows for both sets to have density as low as $(\log N)^{-1+o(1)}$; further, one set may even have exponentially small density $e^{-C(\log N)^{1/5+o(1)}}$.

The analogous problem for three-fold sumsets was first studied by Freiman, Halberstam and Ruzsa [18], who established that the sumset $A+A+A$ contains a much longer progression: indeed of length at least $N^{c\alpha^3}$. Green [29] extended this to $N^{c\alpha^{2+o(1)}}$ and Sanders [78] to $N^{c\alpha^{1+o(1)}}$; however, all of these results required $\alpha \geqslant (\log N)^{-1/2+o(1)}$. In contrast, the best result known for four sets or more, due to Sanders [83], says that the sumset $A+A+A+A$ contains an arithmetic progression of length $N^{c/(\log 2\alpha^{-1})^4}$ when $\alpha \geqslant e^{-C(\log N)^{1/5}}$: in that case all the summands may be rather sparse. In this work we investigate in detail the sumset $A+B+C$, aiming at establishing results valid for sparse sets B and C and in a large range of α.

We now turn to the precise results, starting with the theorem of Croot, Laba and Sisask [9], which constitutes the state-of-the-art on arithmetic progressions in $A+B$.

THEOREM 1.1 (Croot, Laba, Sisask). *Suppose that A and B are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities α and β. Then there exists an absolute constant $c>0$ such that $A+B$ contains an arithmetic progression of length at least*[1]

$$e^{c(\alpha\log N)^{1/2}(\log 2\beta^{-1})^{-3/2}} \quad \textit{if} \quad \alpha\Big(\log\tfrac{\log N}{\beta}\Big)^{-5} \geqslant (c\log N)^{-1}.$$

[1] We assume $N \geqslant 1+\exp(e^e)$ throughout to alleviate logarithmic notation.

In the case of three summands, the best bounds known are due to Sanders [78].

THEOREM 1.2 (Sanders). *Suppose that A, B, C are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities α, β, γ. Then there exists an absolute constant $c > 0$ such that $A + B + C$ contains an arithmetic progression of length at least*

$$N^{c(\alpha\beta\gamma)^{1/3}} \quad \textit{if} \quad (\alpha\beta\gamma)^{1/3} \geqslant (c \log N)^{-1/2} (\log\log N)^{1/2}.$$

Cui, Li and Xue [12] also recently studied the analogous problem for subsets of the primes. We let $\log_k$ denote the logarithm iterated k times below.

THEOREM 1.3 (Cui, Li, Xue). *Suppose that A is a subset of the primes less than N of size $\alpha N/\log N$. Then there exist absolute positive constants c, c_0, c_1 such that $A + A + A$ contains an arithmetic progression of length at least*

$$N^{c\alpha^2/(\log 2\alpha^{-1})} \quad \textit{if} \quad \alpha \geqslant (\log_3 N)^{-c_0},$$
$$N^{c\alpha^4/(\log 2\alpha^{-1})} \quad \textit{if} \quad \alpha \geqslant (\log N)^{-c_1}.$$

Their argument relies on a clever combination of Green's [30] and Helfgott and de Roton's [50] restriction theorems for primes with Green's [29] theorem on $A + A + A$, modified to obtain arithmetic progressions whose elements all have a certain number of representations as a sum of three elements of A. For lack of an existing expression, we call any lower bound on this number of representations a *counting lemma*, here and throughout the article. Motivated by the application to the problem of sumsets of primes, we set out, as a secondary objective, to provide counting lemmas in all our estimates; this is not essentially difficult although it requires some care in the computations.

We now introduce our results. We start with a simple observation which is that the almost-periodicity results of Croot, Laba and Sisask [9] imply a version of Theorem 1.2 which allows for two sets out of three to be sparse, with density as small as $e^{-c(\log N)^{1/5}}$.

THEOREM 1.4. *Suppose that A, B, C are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities α, β, γ. Then there exists an absolute constant $c > 0$ such that $A + B + C$ contains an arithmetic progression of length at least*

$$N^{c\alpha^2/\log^4(2/\alpha\beta\gamma)} \quad \text{if} \quad \alpha\Big(\log \tfrac{2}{\alpha\beta\gamma}\Big)^{-5/2} \geqslant (c \log N)^{-1/2}$$

such that each element of the progression has at least $\frac{1}{2}\alpha\beta\gamma N^2$ representations as a sum $x + y + z$ with $(x, y, z) \in A \times B \times C$.

While the dependency on densities β and γ in Theorem 1.4 is satisfactory, the density α is still required to be at least $(\log N)^{-1/2}$, and the arithmetic progression is shorter than that of Theorem 1.2 when $\alpha = \beta = \gamma$. To overcome these limitations we turn to the argument of Sanders [**78**] to prove Theorem 1.2. The proof there is based on a density-increment strategy, which builds on that introduced by Bourgain [**5**] in the context of Roth's theorem [**69**]. Sanders' recent breakthrough [**81**] in the latter problem introduced very powerful new techniques, and these allow us to revisit the argument of [**78**] so as to obtain the following.

THEOREM 1.5. *Suppose that A, B, C are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities α, β, γ. Then there exists an absolute constant $c > 0$ such that $A + B + C$ contains an arithmetic progression of length at least*

$$N^{c\alpha/\log^5(2/\alpha\beta\gamma)} \quad \text{if} \quad \alpha\Big(\log \tfrac{2}{\alpha\beta\gamma}\Big)^{-7} \geqslant (c \log N)^{-1}$$

such that each element of the progression has at least $e^{-(c\alpha)^{-1}\log^7(2/\alpha\beta\gamma)}\, N^2$ representations as a sum $x + y + z$ with $(x, y, z) \in A \times B \times C$.

Note that the density of each set may now be as low as $(\log N)^{-1+o(1)}$, and that we may take two sets to be very sparse as before. A result of this kind also follows from Theorem 1.1, since an arithmetic progression in $A + B$ is always contained, up to translation, in $A + B + C$; however the arithmetic progression obtained in this way is shorter than the one given by Theorem 1.5, unless γ is

extremely small compared with α and β, for example, when $\alpha \asymp \beta \asymp (\log N)^{-\varepsilon}$ and $\gamma \asymp e^{-C(\log N)^{(1-\varepsilon)/7}}$. Surprisingly, the counting lemma of Theorem 1.5 is quite a lot weaker than that of Theorem 1.4: this is due to the use of an iterative argument which at each step places the sets A, B, C in a certain Bohr set, whose size decreases as we iterate.

By using a generalization by Bloom [2] of the Katz-Koester transform of Sanders [81] to three or more sets, we are able to go one step further in the range of density; however, this time the loss in the counting lemma is substantial.

THEOREM 1.6. *Let $\varepsilon \in (0,1)$ be a parameter and suppose that A, B, C are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities α, β, γ. Then there exists an absolute constant $c > 0$ such that $A + B + C$ contains an arithmetic progression of length at least*

$$\exp\left(c\alpha^{1/4}(\varepsilon \log N)^{1/2}\left(\log \tfrac{2}{\alpha\beta\gamma}\right)^{-7/2}\right) \quad \text{if} \quad \alpha\left(\log \tfrac{2}{\alpha\beta\gamma}\right)^{-14} \geqslant (c\varepsilon \log N)^{-2}$$

such that each element of the progression has at least $N^{2-\varepsilon}$ representations as a sum $x + y + z$ with $(x, y, z) \in A \times B \times C$.

Note that the progression obtained in this way is in fact longer than that of Theorem 1.5 in the range $(\log N)^{-1+o(1)} \leqslant \alpha \leqslant (\log N)^{-2/3+o(1)}$ when, say, $\alpha = \beta = \gamma$ and $\varepsilon \asymp 1$. Finally, we mention two applications of the above results to the analogous problem in the primes. First, since Theorem 1.5 comes with a counting lemma, its conclusion may be inserted into the original argument of Cui, Li and Xue [12] to derive two new estimates, which complement Theorem 1.3.

THEOREM 1.7. *Suppose that A is a subset of the primes less than N of size $\alpha N/\log N$. Then there exist absolute positive constants c, c_2, c_3 such that $A + A + A$ contains an arithmetic progression of length at least*

$$N^{c\alpha/(\log 2\alpha^{-1})^5} \quad \text{if} \quad \alpha \geqslant (\log_4 N)^{-c_2},$$
$$N^{c\alpha^2/(\log 2\alpha^{-1})^5} \quad \text{if} \quad \alpha \geqslant (\log_2 N)^{-c_3}.$$

Secondly, Theorem 1.6, owing to its longer density range, allows us to find long arithmetic progressions in $A+A+A$ for a dense subset A of the primes on grounds of density alone, that is, without appealing to restriction theorems for the primes. This is mostly of conceptual interest, since our argument is also quite involved, relying heavily on methods from [**81**]. We record below the estimate that might be obtained from Theorem 1.6, by observing that the primes have asymptotic density $(\log N)^{-1}$ in the first N integers and with the usual Freiman embedding.

COROLLARY 1.8. *Suppose that A is a subset of the primes less than N of size $\alpha N/\log N$. Then there exists an absolute positive constant c such that $A+A+A$ contains an arithmetic progression of length at least*

$$e^{c(\alpha \log N)^{1/4}(\log\log N)^{-7/2}} \quad \textit{if} \quad \alpha \geqslant (\log N)^{-1}(\log\log N)^{14}.$$

By comparison, the constant c_1 in Theorem 1.3 is $\frac{1}{45}$ in the original argument of [**12**]. The arithmetic progression given by this corollary is, however, shorter than that of Theorems 1.3 and 1.7 in the ranges prescribed there.

We make two last remarks about the shape of the above bounds. The first is that in Theorems 1.4, 1.5 and 1.6, one may assume $\alpha \geqslant \beta \geqslant \gamma$ without loss of generality, and that under this assumption one may replace logarithmic terms $\log \frac{2}{\alpha\beta\gamma}$ by $\log \frac{2}{\beta\gamma}$ there. Secondly, we note that Theorems 1.4–1.7 and Corollary 1.8 are nontrivial if and only if N is larger than an absolute constant.

At this point we should also remark that arithmetic progressions may be obtained for sets much sparser than the ones considered above by a combinatorial method of Croot, Ruzsa and Schoen [**10**], recently generalized in [**44**], although the results there take a rather different form. Indeed, while the Fourier analytic methods used here typically find progressions of length $e^{(\log N)^c}$ in a range of density $\alpha \geqslant (\log N)^{-\delta}$, these combinatorial methods produce shorter progressions, of size $(\log N)^c$, for a larger range of density $\alpha \geqslant N^{-\delta}$.

The article is now organized as follows. Section 2 is devoted to notation and Section 3 is there to recall relevant facts about Bohr sets. The proof of Theorem 1.4 is given in Section 4, and in Section 5 we collect a number of facts on the density-increment strategy which are then used to give the proof of Theorems 1.5 and 1.6 in Section 6. Finally, the estimates of Theorem 1.7 and Corollary 1.8 are derived in Section 7, and comparisons with results on Roth's theorem are drawn in Section 8.

Acknowledgements. We should like to thank our supervisors Régis de la Bretèche and Andrew Granville for discussions that greatly helped improve the exposition in this paper, and we also thank Tom Sanders for many helpful comments.

Funding. This work was supported by a *contrat doctoral* from Université Paris 7.

2. Notation

Here we take a moment to introduce our notation. It is mostly standard up to the choice of normalizations.

General setting. For the rest of the article we fix an integer $N \geqslant 2$ and we write $G = \mathbb{Z}/N\mathbb{Z}$. It is clear, however, that our results are only meaningful when densities vary with N and when N is large: one should think of N as such.

Functions. For a subset X of G and $x \in G$, we define the averaging operator over X, and the operator of translation by x on functions $f : G \to \mathbb{C}$, respectively, by

$$\mathbb{E}_{x\in X} f(x) = \frac{1}{|X|}\sum_{x\in X} f(x) \quad \text{and} \quad \tau_x(f)(u) = f(u+x) \quad \text{for } u \in G.$$

We also occasionally use the identity operator I defined by $If = f$. For any $p \geqslant 1$, we define the L^p-norm of a function f on G by

$$\|f\|_{L^p} = \left(\mathbb{E}_{x\in G}|f(x)|^p\right)^{1/p}.$$

We let $\|f\|_\infty = \sup_{x\in G} |f(x)|$ denote the uniform norm of f over G. The scalar product and the convolution of two functions f, g are defined, respectively, by

$$\langle f, g\rangle_{L^2} = \mathbb{E}_{x\in G} f(x)\overline{g(x)}$$
$$\text{and} \quad f * g(x) = \mathbb{E}_{y\in G} f(y)g(x-y) \qquad (x \in G).$$

We also let $f^{(\ell)} = f * \cdots * f$ denote the convolution of f with itself ℓ times.

Fourier analysis on $\mathbb{Z}/N\mathbb{Z}$. We let $\widehat{G}$ denote the dual group of G, that is, the set of homomorphisms $\gamma : G \to \mathbb{U}$, where $\mathbb{U}$ denotes the unit circle $\{\omega \in \mathbb{C} : |\omega| = 1\}$. We define the Fourier transform $\widehat{f}$ of a function $f : G \to \mathbb{C}$ by

$$\widehat{f}(\gamma) \coloneqq \mathbb{E}_{x\in G} f(x)\overline{\gamma(x)} \qquad (\gamma \in \widehat{G}).$$

The three basic formulæ of Fourier analysis then read as follows:

$$\text{(Inversion)} \qquad f(x) = \textstyle\sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma)\gamma(x),$$
$$\text{(Parseval)} \qquad \langle f, g\rangle_{L^2} = \textstyle\sum_{\gamma\in\widehat{G}} f(\gamma)\overline{g(\gamma)},$$
$$\text{(Convolution)} \qquad \widehat{f * g}(\gamma) = \widehat{f}(\gamma)\widehat{g}(\gamma).$$

For functions $g, h\colon \widehat{G} \to \mathbb{C}$ we also write

$$\|g\|_{\ell^p} = \Big(\sum_{\gamma\in\widehat{G}} |\widehat{g}(\gamma)|^p\Big)^{1/p} \quad \text{and} \quad \langle g, h\rangle_{\ell^2} = \sum_{\gamma\in\widehat{G}} g(\gamma)\overline{h(\gamma)}.$$

Finally, for a real number $\eta > 0$ we define the η-spectrum of a function $f\colon G \to \mathbb{C}$ by

$$\mathrm{Spec}_\eta(f) = \{\gamma \in \widehat{G} \,:\, |\widehat{f}(\gamma)| \geqslant \eta\|f\|_{L^1}\}.$$

Characteristic functions and densities. We let m_G denote the uniform measure on G defined by $m_G(X) = |X|/|G|$ for $X \subset G$. More generally, when A is a subset of G, we let m_A denote the uniform measure on A defined by $m_A(X) = |X \cap A|/|A|$ for $X \subset G$. We also define the normalized characteristic function of a subset A of

G by

$$\mu_A = m_G(A)^{-1} 1_A$$

so that $\|\mu_A\|_{L^1} = 1$; note also the useful identity $1_A * \mu_B(x) = m_{-B}(A - x)$. When B is a subset of G we say that $A \subset B$ has relative density α when $|A| = \alpha|B|$, that is, when $m_B(A) = \alpha$. Note the composition identity $m_G(A) = m_B(A)m_G(B)$.

Asymptotic notation. We let c and C denote absolute positive constants which may take different values at each occurrence. We also make occasional use of Landau's and Vinogradov's asymptotic notation: for two nonnegative functions f and g, we let $f = O(g)$ or $f \ll g$ indicate the fact the $f \leqslant Cg$ for some constant $C > 0$, and $f = \Omega(g)$ or $f \gg g$ indicate that $f \geqslant cg$ for some constant $c > 0$. We write $f \asymp g$ when $f \ll g$ and $f \gg g$.

3. Preliminaries on Bohr sets

Bohr sets are now a standard tool of additive combinatorics. The definition and terminology we use follows Sanders [81, 82]. We also recall the fundamental properties of these sets which will be needed for our work.

DEFINITION 3.1 (Bohr set). *For a set of characters $\Gamma \subset \widehat{G}$ and a real number $\delta > 0$, we let*

$$B(\Gamma, \delta) = \{x \in G : |1 - \gamma(x)| \leqslant \delta \quad \forall \gamma \in \Gamma\}$$

be the Bohr set of frequency set Γ and radius δ. We define $d = |\Gamma|$ to be the dimension of this Bohr set.

Note that $|\gamma(x)| = 1$ and therefore $|1 - \gamma(x)| \leqslant 2$ for every $x \in G$ and $\gamma \in \widehat{G}$, so that the definition is only interesting for $\delta \leqslant 2$. We will often denote a Bohr set simply by the letter B, with associated parameters Γ, δ, d. There is a slight abuse of notation in doing so, as the physical set B may be the same for different frequency sets and radii: one should formally think of B as a triple (B, Γ, δ). We also define

the *dilate* of B by a factor ρ by $B_\rho = B(\Gamma, \delta)_\rho \coloneqq B(\Gamma, \rho\delta)$. Finally we say that B' is a *sub-Bohr set* of B, and we write $B' \leqslant B$, when $\Gamma \subset \Gamma'$ and $\delta' \leqslant \delta$.

We now recall a standard bound on the growth of Bohr sets which is proven in [100, Lemma 4.20], albeit with a slightly different notion of Bohr set. We indicate below the minor changes to the proof needed to recover the following.

LEMMA 3.2 (Doubling ratio of Bohr sets). *Suppose that B is a Bohr set. Then*

$$m_G(B_{1/2}) \geqslant 7^{-d} m_G(B).$$

PROOF. Let $e(x) = e^{2i\pi x}$ and write characters $\gamma : G \to S^1$ as $\gamma = e(\omega)$, where $\omega : G \to \mathbb{R}/\mathbb{Z}$. In [100] a Bohr set of frequency set Γ and radius δ is defined as

$$\tilde{B}(\Gamma, \delta) = \{x : |\omega(x)| \leqslant \delta \ \ \forall \omega \in \Gamma\},$$

whereas here it is defined as

$$B(\Gamma, \delta) = \{x : |1 - e(\omega(x))| \leqslant \delta \ \ \forall \omega \in \Gamma\}.$$

The covering argument used in the proof of [100, Lemma 4.20] may be adjusted via the elementary inclusions

$$\{\omega : |1 - e(\omega)| \leqslant 4\delta\} \subset \{\omega : |\omega| \leqslant \delta\} \subset \{\omega : |1 - e(\omega)| \leqslant 2\pi\delta\},$$

yielding a constant 7 in the final bound in place of 4 there. □

We record an immediate consequence of this bound.

LEMMA 3.3 (Growth of Bohr sets). *Suppose that B is a Bohr set and $\rho \in (0, 1]$. Then*

$$m_G(B_\rho) \geqslant e^{-6d \log 2\rho^{-1}} m_G(B).$$

Observing that $B = B(\Gamma, 2)_{\delta/2}$, this in turn gives the following lemma.

LEMMA 3.4 (Size of Bohr sets). *Suppose that B is a Bohr set of radius $\delta \leqslant 2$. Then*

$$m_G(B) \geqslant e^{-6d \log 4\delta^{-1}}.$$

One essential fact about Bohr sets is that they support a lot of arithmetic structure. A simple illustration of this principle is given by the following easy consequence of Dirichlet's theorem on simultaneous approximation [**85**, Theorem II.1A].

LEMMA 3.5 (Arithmetic progression in a Bohr set). *Let B be a Bohr set of radius $\delta < \pi$. Then B contains an arithmetic progression of size at least $(1/2\pi)\,\delta N^{1/d}$.*

We now recall the notion of regularity of Bohr sets which is of crucial importance for the proof of Theorems 1.5 and 1.6. This is not needed for the proof of Theorem 1.4, therefore the reader only interested in that result may very well skip the following discussion.

Bourgain [**5**] introduced the notion of regular Bohr sets in the context of Roth's theorem. In that situation one often needs to work with Bohr sets on different scales, and it is therefore desirable that the size of dilates $B_{1+\rho}$ vary continuously with ρ.

DEFINITION 3.6 (Regular Bohr set). *Let C_0 be an absolute constant. A Bohr set B is said to be regular for C_0 if*

$$1 - C_0|\rho|d \leqslant \frac{|B_{1+\rho}|}{|B|} \leqslant 1 + C_0|\rho|d \qquad (0 < |\rho| < \tfrac{1}{C_0 d}). \tag{3.1}$$

An essential observation of Bourgain [**5**] is that one may always ensure the regularity of a Bohr set up to dilation by a constant factor.

LEMMA 3.7 (Existence of regular Bohr sets). *There exists an absolute constant C_0 such that for every Bohr set B, there exists $\kappa \in \left[\frac{1}{2}, 1\right)$ such that B_κ is regular for C_0.*

The proof of this result can now be found in many places and we refer, for example, to Proposition 3.5 of [80]. From now on we fix C_0 and we simply say that a Bohr set B satisfying (3.1) is regular. The regularity property allows for a very useful averaging lemma, first formalized by Bourgain as [5, Lemma 3.16]. The version we record below is closest to [35, Lemma 4.2]; it says that Bohr sets are roughly invariant under translation by, or averaging over, elements of a smaller Bohr set.

LEMMA 3.8 (Regularity averaging lemma). *Suppose that B is a regular Bohr set and let $x \in G$ and $\lambda : G \to \mathbb{C}$ with $\|\lambda\|_{L^1} = 1$. Then*

$$\|\mu_{x+B} - \mu_B\|_{L^1} \leqslant C_1 \rho d \qquad \text{if} \quad x \in B_\rho,$$
$$\|\mu_B * \lambda - \mu_B\|_{L^1} \leqslant C_1 \rho d \qquad \text{if} \quad \mathrm{Supp}(\lambda) \subset B_\rho,$$

provided $\rho \leqslant \frac{1}{C_0 d}$ and where $C_1 = 2C_0$.

PROOF. Observe that $\|\mu_{x+B} - \mu_B\|_{L^1} = \frac{1}{|B|} \sum_{y \in G} |1_{x+B}(y) - 1_B(y)|$ and that 1_B and 1_{x+B} are equal on $B_{1-\rho}$ and outside $B_{1+\rho}$. Therefore, $\|\mu_{x+B} - \mu_B\|_{L^1} \leqslant \frac{1}{|B|}(|B_{1+\rho}| - |B_{1-\rho}|)$ and the first bound follows from (3.1). Summing over x with weights $\lambda(x)$ and applying the triangle inequality yields the second estimate. □

4. The Croot-Laba-Sisask approach

The aim of this section is to prove Theorem 1.4. This result is a rather direct consequence of [9, Theorem 7.1] due to Croot, Laba and Sisask, which says that the set of almost-periods of a convolution is guaranteed to contain a large Bohr set. The proof of this theorem relies on a combination of the Croot-Sisask lemma [11] and Chang's spectral lemma [7, Lemmas 3.1 and 3.4]; this combination was first exploited by Sanders [81, 83]. For our purpose we only need the following special case.

LEMMA 4.1 (Bohr-almost-periodicity of convolutions). *Let $p \geqslant 2$ and $\theta \in (0,1)$ be a pair of parameters. Suppose that A_1, A_2 are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities α_1, α_2. Then there exists a Bohr set B such that*

$$\|1_{A_1} * \mu_{A_2} - \tau_x 1_{A_1} * \mu_{A_2}\|_{L^p} \leqslant \theta \alpha_1^{1/p} \qquad (x \in B)$$

with dimension and radius satisfying

$$d \leqslant Cp\theta^{-2}(\log \tfrac{2}{\theta\alpha_1\alpha_2})^3,$$
$$\delta \geqslant c(\theta\alpha_1\alpha_2/p)^C.$$

PROOF. Apply Theorem 7.4 of [9] with $A = A_2$, $B = A_1$, and $S = G$, with doubling constants $K_1 = 2/\alpha_2$ and $K_2 = 2/\alpha_1$, and with $\varepsilon = \theta$. This yields a parameter

$$\delta' = c\theta\alpha_2^{1/2}\alpha_1^{1/p-1/2} \geqslant c\theta\alpha_2^{1/2}$$

and a Bohr set of dimension at most

$$d \leqslant Cp\theta^{-2}(\log 2/\delta')^2(\log 2/\alpha_2) \leqslant Cp\theta^{-2}\Big(\log \tfrac{2}{\theta\alpha_1\alpha_2}\Big)^3$$

and radius

$$\delta = \delta'/d \geqslant cp^{-1}\theta^3\alpha_2^{1/2}\Big(\log \tfrac{2}{\theta\alpha_1\alpha_2}\Big)^{-3} \gg (\theta\alpha_1\alpha_2/p)^4$$

satisfying the desired almost-periodicity property. The bound on δ might seem less crude once we note that the lower bound of Lemma 3.4 on $\log m_G(B)$ depends linearly on d and $\log 2\delta^{-1}$. We have also been somewhat imprecise in handling logarithmic terms, so as not to needlessly clutter the main estimates: indeed these terms have little bearing on the quality of the final results. □

From Lemma 4.1 we first obtain a result slightly more general than Theorem 1.4 which finds a translate of a Bohr set in a sumset. We follow the proof of the similar

Theorem 1.7 on p. 1380 of [11], relying on little more than an elementary identity of convolutions.

PROPOSITION 4.2. *Suppose that A_1, A_2, A_3 are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities $\alpha_1, \alpha_2, \alpha_3$. Then there exists $z \in G$ and a Bohr set B with*

$$d \leqslant C\alpha_1^{-2}\Big(\log \tfrac{2}{\alpha_1\alpha_2\alpha_3}\Big)^4$$
$$\delta \geqslant c(\alpha_1\alpha_2\alpha_3)^C$$

such that $1_{A_1} * 1_{A_2} * 1_{A_3}(y) \geqslant \frac{1}{2}\alpha_1\alpha_2\alpha_3$ for every $y \in z + B$.

PROOF. Apply Lemma 4.1 to A_1 and A_2 with parameters p and θ to be determined later. This yields a Bohr set B with dimension $d \leqslant Cp\theta^{-2}\Big(\log \frac{2}{\theta\alpha_1\alpha_2}\Big)^3$ and radius $\delta \geqslant c(\theta\alpha_1\alpha_2/p)^C$ such that

$$\|(I - \tau_x)1_{A_1} * \mu_{A_2}\|_{L^p} \leqslant \theta\alpha_1^{1/p} \qquad (x \in B). \tag{4.1}$$

Let $z \in G$ and $x \in B$ and observe that

$$1_{A_1} * \mu_{A_2} * \mu_{A_3}(z) - 1_{A_1} * \mu_{A_2} * \mu_{A_3}(z + x) = \langle (I - \tau_x)1_{A_1} * \mu_{A_2}, \tau_{-z}\mu_{-A_3}\rangle_{L^2}.$$

Applying successively Hölder's inequality and (4.1) we have therefore

$$\begin{aligned} |1_{A_1} * \mu_{A_2} * \mu_{A_3}(z) - 1_{A_1} * \mu_{A_2} * \mu_{A_3}(z + x)| &\leqslant \|(I - \tau_x)1_{A_1} * \mu_{A_2}\|_{L^p}\|\mu_{A_3}\|_{L^q} \\ &\leqslant \theta(\alpha_1/\alpha_3)^{1/p} \\ &\leqslant \theta\alpha_3^{-1/p} \end{aligned} \tag{4.2}$$

Since $\mathbb{E}_{z\in G} 1_{A_1} * \mu_{A_2} * \mu_{A_3}(z) = \alpha_1$, we may pick z so that $1_{A_1} * \mu_{A_2} * \mu_{A_3}(z) \geqslant \alpha_1$. Choosing $p = 2 + \log \alpha_3^{-1}$ and $\theta = \alpha_1/2e$, we have $\theta\alpha_3^{-1/p} \leqslant \alpha_1/2$, and by (4.2) we conclude that $1_{A_1} * \mu_{A_2} * \mu_{A_3}(z + x) \geqslant \alpha_1/2$, where $x \in B$ is arbitrary. □

We may now quickly derive Theorem 1.4, which we reproduce below with adjusted notation for convenience.

PROPOSITION (Theorem 1.4). *Suppose that A_1, A_2, A_3 are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities $\alpha_1, \alpha_2, \alpha_3$ and write $\widetilde{\alpha} = \alpha_1\alpha_2\alpha_3$. Then there exist absolute constants $c > 0$ and $C > 0$ such that $A_1 + A_2 + A_3$ contains an arithmetic progression of length at least*

$$N^{c\alpha_1^2/(\log 2\widetilde{\alpha}^{-1})^4} \quad \text{if} \quad \alpha_1(\log 2\widetilde{\alpha}^{-1})^{-5/2} \geqslant C(\log N)^{-1/2}$$

such that each element of the progression has at least $\frac{1}{2}\widetilde{\alpha}N^2$ representations as a sum.

PROOF. Apply Proposition 4.2 to obtain a Bohr set B and an element $z \in G$ such that $d \leqslant C\alpha_1^{-2}(\log 2\widetilde{\alpha}^{-1})^4$, $\delta \geqslant c\widetilde{\alpha}^C$ and $1_{A_1} * 1_{A_2} * 1_{A_3}(y) \geqslant \frac{1}{2}\widetilde{\alpha}$ for every $y \in z + B$. By Lemma 3.5 we may find an arithmetic progression $P \subset B$ of size

$$|P| \geqslant \exp\left(\frac{c\alpha_1^2 \log N}{(\log 2\widetilde{\alpha}^{-1})^4} - C\log 2\widetilde{\alpha}^{-1}\right).$$

Restricting to $\alpha_1^2(\log 2\widetilde{\alpha}^{-1})^{-5} \geqslant C'(\log N)^{-1}$ with C' large enough we see that $z + P$ is the desired arithmetic progression. □

5. Preliminaries on the density-increment strategy

The proof of Theorems 1.5 and 1.6 is based on the density-increment strategy used by Bourgain [**5**, **6**] to obtain good bounds in Roth's theorem [**69**] and later considerably expanded by Sanders in [**81**, **82**]. The base of this theory is best presented in [**83**], while the more advanced techniques specific to Roth's theorem may be found in [**81**, **82**]. We also use a recent refinement of those by Bloom [**2**]. In this section we collect the main facts that we need from these references.

We first need a special case of [**82**, Lemmas 4.6 and 6.3], which together constitute a local version of Chang's spectral lemma [**7**, Lemmas 3.1 and 3.4].

LEMMA 5.1 (Local spectrum annihilation). *Let $\varepsilon \in (0, 1]$ be a parameter. Let B be a regular Bohr set and suppose that $X \subset B$ has relative density τ. Then there*

exists a regular Bohr set $B' \leqslant B$ *with*

$$d' \leqslant d + C\varepsilon^{-2}\log 2\tau^{-1} \quad \textit{and} \quad \delta' \geqslant c\delta/(d^2\varepsilon^{-2}\log 2\tau^{-1})$$

such that $|1-\gamma(x)| \leqslant \frac{1}{2}$ *for every* $\gamma \in \mathrm{Spec}_\varepsilon(\mu_X)$ *and* $x \in B'$.

PROOF. Write $B = B(\Gamma,\delta)$ and let $\Delta = \mathrm{Spec}_\varepsilon(\mu_X)$. By Sanders [82, Lemma 4.6], Δ has $(1,\mu_B)$-relative entropy $k \ll \varepsilon^{-2}\log 2\tau^{-1}$ (see [82] for the definition of this concept); note in passing that, by the definition of entropy, $k \geqslant 1$. Applying [82, Lemma 6.3] to Δ with $\eta = 1$, we may further find a set Λ of size at most k such that, for every $\nu \in (0,1), \rho \leqslant c/(dk)$, and $\gamma \in \Delta$,

$$|1-\gamma(x)| \ll k\nu + \rho d^2(k+1) \quad \text{uniformly in} \quad x \in B(\Gamma \cup \Lambda, \min(\rho\delta, 2\nu)).$$

Choosing $\rho = c/(d^2k)$ and $\nu = c/k$ with c small enough we see that $|1-\gamma(x)| \leqslant \frac{1}{2}$ for $x \in B(\Gamma\cup\Lambda, c\delta/d^2k) =: \widetilde{B}$, and we are done upon choosing $B' = \widetilde{B}_\kappa$ with $\kappa \in \left[\frac{1}{2},1\right)$ chosen via Lemma 3.7 such that $\widetilde{B}$ is regular. □

Note that, as in [78], we need to keep track of the radius of the Bohr set rather than its size, since we are looking for arithmetic progressions such as given by Lemma 3.5. The following is [81, Lemma 3.8] where we used the Bohr set given by Lemma 5.1 in the proof instead. This lemma forms the backbone of the density-increment strategy.

LEMMA 5.2 (L^2 density-increment). *Let* $\nu,\eta,\rho \in (0,1]$ *be parameters. Let* B *and* $\dot{B} \leqslant B_\rho$ *be regular Bohr sets. Suppose that* $A \subset B$ *has relative density* α *and* $X \subset \dot{B}$ *has relative density* τ. *Write* $f_A = 1_A - \alpha 1_B$, *and suppose that* $\rho \leqslant c\nu\alpha/d$ *and*

$$\sum_{\gamma\in\mathrm{Spec}_\eta(\mu_X)} |\widehat{f}_A(\gamma)|^2 \geqslant \nu\alpha^2 m_G(B).$$

Then there exists a regular Bohr set $\breve{B} \leqslant \dot{B}$ *such that* $\|1_A * \mu_{\breve{B}}\|_\infty \geqslant (1+c\nu)\alpha$,

$$\breve{d} \leqslant \dot{d} + C\eta^{-2}\log 2\tau^{-1} \quad \text{and} \quad \breve{\delta} \geqslant c\dot{\delta}/(\dot{d}^2\eta^{-2}\log 2\tau^{-1}).$$

The slightly different shape of the density-increment lemma above affects in a minor way the statement of two results we introduce next. The first is the Katz-Koester transform developed by Sanders [81]; the following is Proposition 4.1 from there.

LEMMA 5.3 (Katz-Koester transform). *Let* $\rho, \rho' \in (0,1)$ *be parameters. Let* B *be a regular Bohr set, assume that* $B' = B_\rho$ *is regular and let* $B'' = B'_{\rho'}$. *Suppose that* $A \subset B$ *has relative density* α *and* $A' \subset B'$ *has relative density* α'. *Assume that* $\rho \leqslant c\alpha/d$ *and* $\rho' \leqslant c\alpha'/d$. *Then either*

(i) *there exists a regular Bohr set* $\breve{B} \leqslant B'$ *such that* $\|1_A * \mu_{\breve{B}}\|_\infty \geqslant (1+c)\alpha$,

$$\breve{d} \leqslant d + C\alpha^{-1}\log 2\alpha'^{-1} \quad \text{and} \quad \breve{\delta} \geqslant c\rho(\alpha\alpha'/d)^C\delta,$$

(ii) *or there exist* $L \subset B$ *with relative density* λ *and* $S \subset B''$ *with relative density* σ, *such that* $\lambda \gg 1$, $\sigma \geqslant e^{-C\alpha^{-1}\log 2\alpha'^{-1}}$ *and*

$$1_L * 1_S \leqslant C\alpha^{-1}\, 1_A * 1_{A'}.$$

A second result we import is a generalization of the above for three of more sets due to Bloom [2]; the following is a direct consequence of the case $k = 2$ of Theorem 6.1 from there.

LEMMA 5.4 (Katz-Koester transform for three sets). *Let* $\rho, \rho' \in (0,1)$ *be parameters. Let* B *be a regular Bohr set, suppose that* $B' = B_\rho$ *is regular and let* $B'' = B'_{\rho'}$. *Suppose that* $A \subset B$ *has relative density* α *and* $A'_1, A'_2 \subset B'$ *have relative densities* α'_1, α'_2, *and write* $\gamma = \alpha\alpha'_1\alpha'_2$. *Assume that* $\rho \leqslant c\alpha/d$ *and* $\rho' \leqslant c\gamma/d$. *Then either*

(i) *there exists a regular Bohr set* $\breve{B} \leqslant B'$ *such that* $\|1_A * \mu_{\breve{B}}\|_\infty \geqslant (1+c)\alpha$,

$$\breve{d} \leqslant d + C\alpha^{-1/2}\log 2\gamma^{-1} \quad \text{and} \quad \breve{\delta} \geqslant c\rho(\gamma/d)^C\delta,$$

(ii) *or there exist* $L \subset B$ *with relative density* λ *and* $S_1, S_2 \subset B''$ *with relative densities* σ_1, σ_2 *such that* $\lambda \gg 1$, $\sigma_i \geqslant e^{-C\alpha^{-1/2}\log 2\gamma^{-1}}$, *and*

$$1_L * 1_{S_1} * 1_{S_2} \leqslant C\alpha^{-2}\, 1_A * 1_{A_1'} * 1_{A_2'}.$$

Finally, we are going to make extensive use of the Croot-Sisask lemma [**11**], which says that two-fold convolutions possess large sets of almost-periods. This technique is particularly suited to prove asymmetric results such as Theorems 1.5 and 1.6. The slightly different version we quote is [**83**, Lemma 4.3] due to Sanders.

LEMMA 5.5 (Croot-Sisask lemma). *Let* $p \geqslant 2$ *and* $\varepsilon \in (0,1)$ *be a pair of parameters. Let* $f : G \to \mathbb{C}$ *and* $L \geqslant 1$ *and assume that* S *and* T *are subsets of* G *such that* $|S+T| \leqslant L|S|$. *Then there exist* $t \in T$ *and a set* $X \subset T$ *of size* $|X| \geqslant (2L)^{-Cp/\varepsilon^2}|T|$ *such that*

$$\|f * \mu_S - \tau_y f * \mu_S\|_{L^p} \leqslant \varepsilon \|f\|_{L^p} \qquad (y \in X - t).$$

This has the following familiar consequence, often used implicitly throughout the literature.

LEMMA 5.6 (L^p-smoothing of convolutions). *Let* $p \geqslant 2$, $\ell \geqslant 1$, *and* $\theta \in (0,1)$ *be parameters. Let* $f : G \to \mathbb{C}$ *and* $L \geqslant 1$ *and suppose that* S *and* T *are subsets of* G *such that* $|S+T| \leqslant L|S|$. *Then there exists a set* $X \subset T$ *of size* $|X| \geqslant (2L)^{-Cp\ell^2/\theta^2}|T|$ *such that*

$$\|f * \mu_S - f * \mu_S * \lambda_X^{(\ell)}\|_{L^p} \leqslant \theta \|f\|_{L^p}$$

where $\lambda_X = \mu_X * \mu_{-X}$.

PROOF. Apply Lemma 5.5 with parameter $\varepsilon = \theta/(2\ell)$. By the triangle inequality and the translation invariance of L^p-norms, we have, for every $x_1, \dots, x_\ell, x'_1, \dots, x'_\ell \in X$:

$$\|f * \mu_S - \tau_{x_1 - x'_1 + \dots + x_\ell - x'_\ell} f * \mu_S\|_{L^p} \leqslant \theta \|f\|_{L^p}.$$

By averaging over the numerous x_i, x'_j and the triangle inequality we recover the result. □

6. Proof of Theorems 1.5 and 1.6

We are now ready to start with the proof of our main estimates. In this section we introduce a new piece of notation to make computations more bearable: to every Bohr set B we associate the *density parameter* $b = m_G(B)$. We start with an easy consequence of regularity that gives us some control on the size of scaled-down sets.

LEMMA 6.1 (Scaling lemma). *Let $\rho \in (0,1)$ be a parameter. Let B be a regular Bohr set and $B' \subset B_\rho$. Suppose that $A \subset B$ has relative density α and $\rho \leqslant c/d$, then*

$$\|1_A * \mu_{B'}\|_\infty \geqslant (1 - O(\tfrac{\rho d}{\alpha}))\alpha.$$

PROOF. We have, by Lemma 3.8,

$$\begin{aligned}
\mathbb{E}_{x \in B} 1_A * \mu_{B'}(x) &= \langle 1_A * \mu_{B'}, \mu_B \rangle_{L^2} \\
&= \langle 1_A, \mu_B * \mu_{B'} \rangle_{L^2} \\
&= \langle 1_A, \mu_B \rangle_{L^2} + O\Big(\|\mu_B - \mu_B * \mu_{B'}\|_{L^1} \|1_A\|_\infty \Big) \\
&= \alpha + O(\rho d).
\end{aligned}$$

Bounding the left-hand side in $\|\cdot\|_\infty$ norm concludes the proof. □

Our iterative argument initially follows that developed by Sanders in [78], with slight modifications to accommodate upper level sets. We recall its principle here.

At each step, one fixes a small Bohr set B' and finds a translate A_3' of A_3 with relative density in B' of same order as that of A_3 in B. Then either B' is contained in the upper level set $\{1_{A_1} * 1_{A_2} * 1_{A_3'} > K\}$, or it has nonempty intersection $\mathcal{U}$ with the lower level set $\{1_{A_1} * 1_{A_2} * 1_{A_3'} \leqslant K\}$. The scalar product $\langle 1_{A_1} * 1_{A_2} * 1_{A_3'}, 1_{\mathcal{U}} \rangle_{L^2}$ is then unusually small for a good choice of K. The usual density-increment strategy then allows one to find a smaller Bohr set on which either A_1 or A_2 has increased density. Since the density is bounded by 1, we may iterate this process only a finite number of times, after which we have found a translate of a Bohr set in a certain upper level set.

At this point, however, we take advantage of two techniques from [**81**], which we apply in a similar fashion. The first is the Katz-Koester transform which in this situation roughly redistributes the mass of the sets A_1 and A_3' on two new sets L and S where L is thick and S is not too small, without affecting the size of the convolution $1_{A_1} * 1_{A_3'}$ excessively. The second is the Croot-Sisask lemma which allows one to smooth the convolution $1_L * 1_S$ by a factor $\lambda_X^{(\ell)}$. At last the density-increment strategy makes it possible to exploit the smallness of the new scalar product $\langle 1_L * 1_S * 1_{A_2} * \lambda_X^{(\ell)}, 1_{\mathcal{U}} \rangle$ to obtain a density increment on A_2.

Our main iterative lemma is then the following. On a first reading the reader may wish to take $\omega = 0$ below for simplicity, which suffices to obtain Theorem 1.5 without a counting lemma.

PROPOSITION 6.2 (Main iterative lemma). *Let $\rho, \omega \in (0,1)$ be parameters. Let B be a regular Bohr set and suppose that $B' = B_\rho$ is regular. Suppose that $A_1, A_2, A_3 \subset B$ have relative densities $\alpha_1, \alpha_2, \alpha_3$ and write $\widetilde{\alpha} = \alpha_1\alpha_2\alpha_3$. Assume that $\rho \leqslant c\widetilde{\alpha}/d$ and $\omega \leqslant e^{-C(d+\alpha_1^{-1})\log(2d/\rho\widetilde{\alpha})}$. Then either*

(i) *there exists a regular Bohr set* $\breve{B} \leqslant B$ *such that, for some* $i \in \{1,2\}$,

$$\begin{aligned} \|1_{A_i} * \mu_{\breve{B}}\|_\infty &\geqslant (1+c)\alpha_i, \\ \breve{d} &\leqslant d + C\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^4, \\ \breve{\delta} &\geqslant c\rho(\widetilde{\alpha}/d)^C\delta, \end{aligned}$$

(ii) *or there exists* $x \in G$ *such that* $B' \subset \{\, y : 1_{A_1} * 1_{A_2} * 1_{A_3}(x+y) > \omega b^2 \,\}$.

PROOF. By Lemma 6.1 we may find $x \in G$ such that $A_3' = (A_3 - x) \cap B'$ has relative density in B' equal to $\alpha_3' = 1_{A_3} * \mu_{B'}(x) \gg \alpha_3$. Now define

$$\mathcal{U} = \{\, y : 1_{A_1} * 1_{A_2} * 1_{A_3}(x+y) \leqslant \omega b^2 \,\} \cap B',$$

we may assume that $\mathcal{U}$ is nonempty since else we are in the second case of the proposition. Note that from the inclusion $A_3' \subset A_3 - x$ and the definition of $\mathcal{U}$, we have

$$\begin{aligned} \langle 1_{A_1} * 1_{A_2} * 1_{A_3'}, \mu_{\mathcal{U}} \rangle_{L^2} &\leqslant \langle 1_{A_1} * 1_{A_2} * 1_{A_3 - x}, \mu_{\mathcal{U}} \rangle_{L^2} \\ &= \langle 1_{A_1} * 1_{A_2} * 1_{A_3}, \mu_{x+\mathcal{U}} \rangle_{L^2} \\ &\leqslant \omega b^2 \end{aligned} \tag{6.1}$$

where $\mu_{\mathcal{U}}$ is well-defined since $\mathcal{U} \neq \varnothing$. From hereon, the proof divides into three steps.

Applying the Katz-Koester transform. Let $\rho' = c\kappa\alpha_3/d$ and $B'' = B'_{\rho'}$, where $\kappa \in \left[\frac{1}{2}, 1\right)$ is chosen via Lemma 3.7 so that B'' is regular. Applying Lemma 5.3 to $A = A_1$ and $A' = A_3'$ with parameters ρ and ρ' then results in one of two cases. In case (i) of that lemma we obtain a regular Bohr set $\breve{B} \leqslant B'$ such that $\|1_{A_1} * \mu_{\breve{B}}\|_\infty \geqslant (1+c)\alpha_1$,

$$\breve{d} \leqslant d + C\alpha_1^{-1}\log 2\alpha_3^{-1} \quad \text{and} \quad \breve{\delta} \geqslant c\rho(\alpha_1\alpha_3/d)^C\delta,$$

which is enough to conclude. In case (ii), we may find $L \subset B$ with relative density λ and $S \subset B''$ with relative density σ such that

$$\lambda \gg 1 \quad \text{and} \quad \sigma \geqslant e^{-C\alpha_1^{-1}\log 2\alpha_3^{-1}}, \tag{6.2}$$

$$1_L * 1_S \ll \alpha_1^{-1} 1_{A_1} * 1_{A_3'}. \tag{6.3}$$

By (6.3) we then have

$$\begin{aligned}\langle 1_L * \mu_S, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2} &= (\sigma b'')^{-1}\langle 1_L * 1_S, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2} \\ &\ll (\alpha_1\sigma b'')^{-1}\langle 1_{A_1} * 1_{A_3'}, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2} \\ &= (\alpha_1\sigma b'')^{-1}\langle 1_{A_1} * 1_{A_2} * 1_{A_3'}, \mu_{\mathcal{U}}\rangle_{L^2}.\end{aligned}$$

By (6.1) we have further

$$\begin{aligned}\langle 1_L * \mu_S, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2} &\ll (\alpha_1\sigma b'')^{-1}\omega b^2 \\ &= (\lambda\alpha_1\alpha_2\sigma)^{-1}(b/b'')\omega \cdot \lambda\alpha_2 b.\end{aligned}$$

Recalling (6.2) and applying Lemma 3.3 we have therefore

$$\langle 1_L * \mu_S, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2} \leqslant e^{C(d+\alpha_1^{-1})\log(2d/\rho\widetilde{\alpha})}\omega \cdot \lambda\alpha_2 b.$$

Assuming $\omega \leqslant e^{-C'(d+\alpha_1^{-1})\log(2d/\rho\widetilde{\alpha})}$ with C' large enough we eventually obtain

$$\langle 1_L * \mu_S, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2} \leqslant \tfrac{1}{4}\lambda\alpha_2 b. \tag{6.4}$$

Applying the Croot-Sisask lemma. Let $\rho'' = c\kappa'/d$ and $B''' = B''_{\rho''}$, where $\kappa' \in \left[\frac{1}{2}, 1\right)$ is chosen via Lemma 3.7 so that B''' is regular, and with c small enough so that, by regularity of B'' and Definition 3.6,

$$|S + B'''| \leqslant |B'' + B'''| \leqslant |B''_{1+\rho''}| \leqslant 2|B''| = (2/\sigma)|S|.$$

Applying Lemma 5.6 to $f = 1_L$ and $T = B'''$ with parameters p, ℓ, θ to be determined later, we obtain a set $X \subset B'''$ of relative density τ with

$$\tau \geqslant \exp\Big(-C(p\ell^2/\theta^2)\log 2\sigma^{-1}\Big) \tag{6.5}$$

such that

$$\|1_L * \mu_S - 1_L * \mu_S * \lambda_X^{(\ell)}\|_{L^p} \leqslant \theta \|1_L\|_{L^p}.$$

By Hölder's and Young's inequalities we have therefore

$$\begin{aligned}
|\langle 1_L * \mu_S, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2} - \langle 1_L * \mu_S * \lambda_X^{(\ell)}, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2}| & \\
&\leqslant \|1_L * \mu_S - 1_L * \mu_S * \lambda_X^{(\ell)}\|_{L^p} \|1_{-A_2} * \mu_{\mathcal{U}}\|_{L^q} \\
&\leqslant \theta \|1_L\|_{L^p} \|1_{-A_2}\|_{L^q} \\
&= \theta \lambda^{1/p} \alpha_2^{1-1/p} b
\end{aligned}$$

Choosing $p = 2 + \log \alpha_2^{-1}$ and $\theta = \lambda^{1-1/p}/4e \asymp 1$, this is less than $\frac{1}{4}\lambda\alpha_2 b$, which combined with (6.4) shows that

$$|\langle 1_L * \mu_S * \lambda_X^{(\ell)}, 1_{-A_2} * \mu_{\mathcal{U}}\rangle_{L^2}| \leqslant \tfrac{1}{2}\lambda\alpha_2 b. \tag{6.6}$$

Obtaining an L^2 density increment. Since $\mathcal{U}, S, X$ are contained in B', the function $\mu_{\mathcal{U}} * \mu_{-S} * \lambda_X^{(\ell)}$ has support in $(2\ell+2)B' \subset B_{(2\ell+2)\rho}$ and we have, by Lemma 3.8,

$$\begin{aligned}
\langle 1_L * \mu_S * \lambda_X^{(\ell)}, 1_B * \mu_{\mathcal{U}}\rangle_{L^2} &= \langle 1_L, 1_B * \mu_{\mathcal{U}} * \mu_{-S} * \lambda_X^{(\ell)}\rangle_{L^2} \\
&= \langle 1_L, 1_B\rangle_{L^2} + O\Big(\|1_B - 1_B * \mu_{\mathcal{U}} * \mu_{-S} * \lambda_X^{(\ell)}\|_{L^1} \|1_L\|_\infty\Big) \\
&= \lambda b + O(\ell\rho d b) \\
&\geqslant \tfrac{3}{4}\lambda b
\end{aligned} \tag{6.7}$$

provided that $\rho \leqslant c/(\ell d)$, which will turn out to be the case. Forming the balanced function $f_{-A_2} = 1_{-A_2} - \alpha_2 1_B$, we deduce from (6.6) and (6.7) that

$$|\langle 1_L * \mu_S * \lambda_X^{(\ell)}, f_{-A_2} * \mu_{\mathcal{U}} \rangle_{L^2}| \geqslant \tfrac{1}{4}\lambda\alpha_2 b.$$

By Parseval's formula and the inequality $\|\widehat{f}\|_\infty \leqslant \|f\|_{L^1}$ we have therefore

$$\begin{aligned}\tfrac{1}{4}\lambda\alpha_2 b &\leqslant \left|\langle \widehat{1}_L \cdot \widehat{\mu}_S \cdot \widehat{\mu}_X^\ell \cdot \widehat{\mu}_{-X}^\ell, \widehat{f}_{-A_2} \cdot \widehat{\mu}_{\mathcal{U}} \rangle_{\ell^2}\right| \\ &\leqslant \|\widehat{\mu}_S\|_\infty \|\widehat{\mu}_{\mathcal{U}}\|_\infty \|\widehat{1}_L \cdot \widehat{f}_{A_2} \cdot \widehat{\mu}_X^{2\ell}\|_{\ell^1} \\ &\leqslant \|\widehat{1}_L \cdot \widehat{f}_{A_2} \cdot \widehat{\mu}_X^{2\ell}\|_{\ell^1}.\end{aligned}$$

By Cauchy-Schwarz and Parseval's identity, we then have

$$\tfrac{1}{4}\lambda\alpha_2 b \leqslant \|\widehat{1}_L\|_{\ell^2} \|\widehat{f}_{A_2} \cdot \widehat{\mu}_X^{2\ell}\|_{\ell^2} = (\lambda b)^{1/2} \|\widehat{f}_{A_2} \cdot \widehat{\mu}_X^{2\ell}\|_{\ell^2}.$$

It follows that, for some constant c,

$$\sum_\gamma |\widehat{f}_{A_2}(\gamma)|^2 |\widehat{\mu}_X(\gamma)|^{4\ell} \geqslant \tfrac{1}{16}\lambda\alpha_2^2 b \geqslant c\alpha_2^2 b. \tag{6.8}$$

By Parseval's identity and choosing $\ell = C \log 2\alpha_2^{-1}$ with C large enough we have

$$\begin{aligned}\sum_{\gamma\,:\,|\widehat{\mu}_X(\gamma)| \leqslant 1/2} |\widehat{f}_{A_2}(\gamma)|^2 |\widehat{\mu}_X(\gamma)|^{4\ell} &\leqslant 2^{-4\ell} \|f_{A_2}\|_{L^2}^2 \\ &\leqslant 2^{2-4\ell}\alpha_2 b \\ &\leqslant \tfrac{1}{2}c\alpha_2^2 b.\end{aligned}$$

By (6.8) and the bound $\|\widehat{\mu}_X\|_\infty \leqslant 1$, we have therefore

$$\sum_{\gamma \in \mathrm{Spec}_{1/2}(\mu_X)} |\widehat{f}_{A_2}(\gamma)|^2 \gg \alpha_2^2 b.$$

The parameters we have chosen have size $p \asymp \log 2\alpha_2^{-1}$, $\ell \asymp \log 2\alpha_2^{-1}$, and $\theta \asymp 1$, and therefore by (6.5) and (6.2), we have

$$\tau \geqslant \exp\Big(-C\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^4\Big).$$

Since $\rho' \asymp \alpha_3/d$ and $\rho'' \asymp 1/d$, we also have $\delta''' = c\rho(\alpha_3/d^2)\delta$. Applying Lemma 5.2 with $A = A_2$ and for $\eta = 1/2$ and some $\nu \asymp 1$, we therefore obtain a regular Bohr set $\breve{B} \leqslant B'''$ such that $\|1_{A_2} * \mu_{\breve{B}}\|_\infty \geqslant (1+c)\alpha_2$ and

$$\breve{d} \leqslant d + C\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^4 \quad \text{and} \quad \breve{\delta} \geqslant c\rho(\widetilde{\alpha}/d)^4\delta,$$

which again is enough to conclude. □

We are now in a position to prove the following result, which gives slightly more structure than Theorem 1.5 in the form of a translate of a large Bohr set. Theorem 1.5 will then follow quickly from this proposition and Lemma 3.5.

PROPOSITION 6.3. *Suppose that A_1, A_2, A_3 are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities $\alpha_1, \alpha_2, \alpha_3$ and write $\widetilde{\alpha} = \alpha_1\alpha_2\alpha_3$. Then there exist $z \in G$ and a Bohr set B with*

$$d \leqslant C\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^5,$$
$$\delta \geqslant \exp\Big(-C(\log 2\widetilde{\alpha}^{-1})^2\Big),$$

such that, for every $y \in z + B$,

$$1_{A_1} * 1_{A_2} * 1_{A_3}(y) > \exp\Big(-C\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^7\Big).$$

PROOF. The proof proceeds by iteration of Proposition 6.2. We construct iteratively a sequence of regular Bohr sets $B^{(i)}$ and sequences of sets $A_1^{(i)}$, $A_2^{(i)}$, $A_3^{(i)} \subset B^{(i)}$ of relative densities $\alpha_1^{(i)}$, $\alpha_2^{(i)}$, $\alpha_3^{(i)}$. We initiate the iteration with $B^{(1)} = B(\{0\}, 2) = \mathbb{Z}/N\mathbb{Z}$, which is regular, and with $(A_1^{(1)}, A_2^{(1)}, A_3^{(1)}) = (A_1, A_2, A_3)$. We denote by

δ_i, d_i, and b_i, respectively the radius, dimension, and density in G of $B^{(i)}$, and we write $\widetilde{\alpha}^{(i)} = \alpha_1^{(i)}\alpha_2^{(i)}\alpha_3^{(i)}$.

At each step i, we apply Proposition 6.2 to the sets $A_1^{(i)}, A_2^{(i)}, A_3^{(i)}$ with parameters ω_i and ρ_i to be determined later. In case (i) of that proposition we define $B^{(i+1)} = \breve{B}^{(i)}$, while in case (ii) we stop the iteration. Whenever $B^{(i+1)}$ is defined we pick $(x_{j,i})_{1\leqslant j\leqslant 3}$ so that, for every j, $A_j^{(i+1)} := (A_j^{(i)} - x_{j,i}) \cap B^{(i+1)}$ has relative density in $B^{(i+1)}$ equal to

$$\alpha_j^{(i+1)} = 1_{A_j^{(i)}} * \mu_{B^{(i+1)}}(x_{j,i}) = \|1_{A_j^{(i)}} * \mu_{B^{(i+1)}}\|_\infty.$$

We now assume that $B^{(i)}$ is defined for $1 \leqslant i \leqslant n$. Let $i < n$, our application of Proposition 6.2 then shows that there exists $j_i \in \{1,2\}$ such that $\alpha_{j_i}^{(i+1)} \geqslant (1+c)\alpha_{j_i}^{(i)}$. Choose now $\rho_i = c'\kappa_i\widetilde{\alpha}^{(i)}/(2i^2 d_i)$, where $\kappa_i \in \left[\frac{1}{2}, 1\right)$ is picked via Lemma 3.7 so that $B_{\rho_i}^{(i)}$ is regular, and with c' small enough so that, by Lemma 6.1,

$$\alpha_j^{(i+1)} \geqslant \Big(1 - O(\rho_i d_i/\alpha_j^{(i)})\Big)\alpha_j^{(i)} \geqslant (1 - \tfrac{c}{2i^2})\alpha_j^{(i)} \tag{6.9}$$

for every $1 \leqslant j \leqslant 3$. This implies that

$$\alpha_1^{(i+1)}\alpha_2^{(i+1)} \geqslant (1 - c/2)(1+c)\alpha_1^{(i)}\alpha_2^{(i)} \geqslant (1 + c/4)\alpha_1^{(i)}\alpha_2^{(i)},$$

and as a consequence the iteration proceeds for at most $n = O(\log 2\widetilde{\alpha}^{-1})$ steps. Iterating (6.9) we also obtain

$$\alpha_j^{(i)} \geqslant e^{-O(\sum_{i=1}^{\infty} i^{-2})}\alpha_j \gg \alpha_j$$

uniformly in $1 \leqslant j \leqslant 3$ and $1 \leqslant i \leqslant n$. The dimension bound from Proposition 6.2 then becomes

$$d_{i+1} \leqslant d_i + (C/\alpha_1^{(i)})\log^4(2/\widetilde{\alpha}^{(i)}) \leqslant d_i + O\Big(\alpha_1^{-1}\log^4(2/\widetilde{\alpha})\Big)$$

for $i < n$ and therefore $d_i \ll i\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^4 \ll \alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^5$ uniformly in $1 \leqslant i \leqslant n$. The radius bound from Proposition 6.2 is then

$$\delta_{i+1} \geqslant (\widetilde{\alpha}^{(i)}/2id_i)^{O(1)}\delta_i \geqslant (\widetilde{\alpha}/2)^{O(1)}\delta_i$$

for $i < n$, whence $\delta_i \geqslant (\widetilde{\alpha}/2)^{O(i)} \geqslant e^{-O((\log 2\widetilde{\alpha}^{-1})^2)}$ uniformly in $1 \leqslant i \leqslant n$.

Finally, we choose $\omega_i = \omega$ independent of i so as to satisfy the condition

$$\omega \leqslant \exp\Big(-C(d_i + (\alpha_1^{(i)})^{-1})\log(2d_i/\rho_i\widetilde{\alpha}^{(i)})\Big)$$

from Proposition 6.2 for every $1 \leqslant i \leqslant n$. From the previous dimension and radius bounds we see that it is enough to take $\omega = e^{-C'\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^6}$, with C' large enough. For that choice we deduce from Lemma 3.4 and the bounds on d_i and δ_i that $\omega b_i{}^2 \geqslant e^{-O(\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^7)}$ uniformly in $1 \leqslant i \leqslant n$. When we are in case (ii) of Proposition 6.2 we therefore find that $B_{\rho_n}^{(n)}$ is contained in a translate of

$$\{\, y \,:\, 1_{A_1^{(n)}} * 1_{A_2^{(n)}} * 1_{A_3^{(n)}}(y) \geqslant \exp\Big(-C\alpha_1^{-1}(\log 2\widetilde{\alpha}^{-1})^7\Big)\,\}.$$

Since $\rho_n \geqslant (\widetilde{\alpha}/2)^{O(1)}$ and the $A_j^{(n)}$ are, by construction, contained in translates of the A_j, this concludes the proof. □

PROOF OF THEOREM 1.5. Applying Proposition 6.3 with $(A_1, A_2, A_3) = (A, B, C)$ and using Lemma 3.5 we may find an arithmetic progression P such that

$$|P| \geqslant \exp\left(\frac{c\alpha(\log N)}{(\log(2/\alpha\beta\gamma))^5} - C(\log(2/\alpha\beta\gamma))^2\right)$$

and an element $z \in G$ such that $1_{A_1} * 1_{A_2} * 1_{A_3}(y) \geqslant e^{-C\alpha^{-1}\log^7(2/\alpha\beta\gamma)}$ for all $y \in z+P$. Restricting to $\alpha(\log\frac{2}{\alpha\beta\gamma})^{-7} \geqslant C'(\log N)^{-1}$ with C' large enough we see that $z + P$ is the desired arithmetic progression. □

We now turn to the slightly more difficult proof of Theorem 1.6. The main strategy is the same and we again start with a small scalar product $\langle 1_{A_1} * 1_{A_3'} * 1_{-\mathcal{U}}, 1_{-A_2}\rangle$

where $\mathcal{U}$ is a certain lower level set. However, we now fully exploit the set $\mathcal{U}$ in applying the generalized Katz-Koester transform from [2] to the three sets $A_1, A_3', -\mathcal{U}$. This redistributes the mass more efficiently and accounts for the improved dependency on densities. The rest of the proof runs similarly with applications of the Croot-Sisask lemma and the density-increment strategy.

This, however, requires us to assume that $\mathcal{U} = \{1_{A_1} * 1_{A_2} * 1_{A_3'} \leqslant K\}$ is dense enough inside a Bohr set B'. We are then in a situation already encountered in [78] where at each step of the iteration it either happens that $\mathcal{U}$ has low density and that the upper level set $\mathcal{U}^c = \{1_{A_1} * 1_{A_2} * 1_{A_3'} > K\}$ is thick inside B'; or that a density increment can be obtained. The following lemma makes this precise and the reader may again let $\omega = 0$ there to obtain Theorem 1.6 without a counting lemma.

PROPOSITION 6.4 (Main iterative lemma). *Let $\rho, v, \omega \in (0,1)$ be parameters. Let B be a regular Bohr set and assume that $B' = B_\rho$ is regular. Suppose that $A_1, A_2, A_3 \subset B$ have relative densities $\alpha_1, \alpha_2, \alpha_3$ and write $\widetilde{\alpha} = \alpha_1\alpha_2\alpha_3$. Assume that $\rho \leqslant c\widetilde{\alpha}/d$ and $w \leqslant e^{-C(d+\alpha_1^{-1/2})\log(2d/\rho v\widetilde{\alpha})}$. Then either*

(i) *there exists a regular Bohr set $\breve{B} \leqslant B$ and $i \in \{1,2\}$ such that*

$$\begin{aligned} \|1_{A_i} * \mu_{\breve{B}}\|_\infty &\geqslant (1+c)\alpha_i, \\ \breve{d} &\leqslant d + C\alpha_1^{-1/2}(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^4, \\ \breve{\delta} &\geqslant c\rho(v\widetilde{\alpha}/d)^C\delta, \end{aligned}$$

(ii) *or there exists $x \in G$ such that $\{y : 1_{A_1} * 1_{A_2} * 1_{A_3}(x+y) > \omega b^2\} \cap B'$ has relative density at least $1 - v$ in B'.*

PROOF. The proof is in many aspects similar to that of Proposition 6.2 and therefore we are more brief in computations. By Lemma 6.1 we may find $x \in G$ such that $A_3' = (A_3 - x) \cap B'$ has relative density $\alpha_3' = 1_{A_3} * \mu_{B'}(x) \gg \alpha_3$ in B'. Let

$$\mathcal{U} = \{y : 1_{A_1} * 1_{A_2} * 1_{A_3}(x+y) \leqslant \omega b^2\} \cap B'$$

have density u in B'; we may assume that $u \geqslant v$ since else we are in the second case of the proposition. Note that, by the definitions of A_3' and $\mathcal{U}$, we have

$$\langle 1_{A_1} * 1_{A_2} * 1_{A_3'}, 1_{\mathcal{U}} \rangle_{L^2} \leqslant \omega b^2 \cdot ub' \leqslant \omega b^2 b'. \tag{6.10}$$

From here the proof again divides into three steps.

Applying the Katz-Koester transform. Choose $\rho' = cv\widetilde{\alpha}/d$ with the help of Lemma 3.7 so that $B'' := B'_{\rho'}$ is regular. Applying Lemma 5.4 with $(A, A_1', A_2') = (A_1, -\mathcal{U}, A_3')$ then results into one of two cases. In case (i) of that lemma we obtain a regular Bohr set $\breve{B} \leqslant B'$ such that $\|1_{A_1} * \mu_{\breve{B}}\|_\infty \geqslant (1+c)\alpha_1$,

$$\breve{d} \leqslant d + C\alpha_1^{-1/2}\log(2/v\widetilde{\alpha}) \quad \text{and} \quad \breve{\delta} \geqslant c\rho(v\widetilde{\alpha}/d)^C\delta,$$

which is enough to conclude via the crude bound $\log(2/v\widetilde{\alpha}) \ll (\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})$. In case (ii), we may find $L \subset B$ of relative density λ and $S_1, S_2 \subset B''$ of relative densities σ_1, σ_2 such that

$$\lambda \gg 1, \quad \sigma_1, \sigma_2 \geqslant e^{-C\alpha_1^{-1/2}\log(2/v\widetilde{\alpha})}, \tag{6.11}$$

$$1_L * 1_{S_1} * 1_{S_2} \ll \alpha_1^{-2}\, 1_{A_1} * 1_{-\mathcal{U}} * 1_{A_3'}. \tag{6.12}$$

In that case write $I = \langle 1_L * \mu_{S_1} * \mu_{S_2}, 1_{-A_2} \rangle_{L^2}$ for convenience. By (6.12) we then have

$$\begin{aligned} I &\ll (\alpha_1^2\sigma_1\sigma_2)^{-1}(b'')^{-2}\langle 1_{A_1} * 1_{-\mathcal{U}} * 1_{A_3'}, 1_{-A_2}\rangle_{L^2} \\ &= (\alpha_1^2\sigma_1\sigma_2)^{-1}(b'')^{-2}\langle 1_{A_1} * 1_{A_2} * 1_{A_3'}, 1_{\mathcal{U}}\rangle_{L^2}. \end{aligned}$$

By (6.10), (6.11) and Lemma 3.3 we have further

$$\begin{aligned} I &\ll (\alpha_1^2\sigma_1\sigma_2)^{-1}(b'')^{-2}\omega b^2 b' \\ &= (\lambda\alpha_1^2\alpha_2\sigma_1\sigma_2)^{-1}(b/b'')(b'/b'')\omega \cdot \lambda\alpha_2 b \\ &\leqslant e^{C(d+\alpha_1^{-1/2})\log(2d/\rho v\widetilde{\alpha})}\omega \cdot \lambda\alpha_2 b. \end{aligned}$$

Assuming $\omega \leqslant e^{-C'(d+\alpha_1^{-1/2})\log(2d/\rho v\widetilde{\alpha})}$ with C' large enough we have therefore

$$\langle 1_L * \mu_{S_1}, 1_{-A_2} * \mu_{-S_2}\rangle_{L^2} = I \leqslant \tfrac{1}{4}\lambda\alpha_2 b. \tag{6.13}$$

Applying the Croot-Sisask lemma. We let $B''' = B''_{\rho''}$ with $\rho'' = c/d$ chosen such that B''' is regular (via Lemma 3.7) and with c small enough so that, by the regularity of B'', $|S_1 + B'''| \leqslant |B''_{1+\rho''}| \leqslant (2/\sigma_1)|S_1|$. Applying Lemma 5.6 with $f = 1_L$, $S = S_1$, $T = B'''$ and parameters p, ℓ, θ to be determined later, we obtain a set $X \subset B'''$ of relative density τ with

$$\tau \geqslant \exp\left(-C(p\ell^2/\theta^2)\log 2\sigma_1^{-1}\right) \tag{6.14}$$

such that

$$\|1_L * \mu_{S_1} - 1_L * \mu_{S_1} * \lambda_X^{(\ell)}\|_{L^p} \leqslant \theta\|1_L\|_{L^p}.$$

Proceeding exactly as in the proof of Proposition 6.2 we then obtain from (6.13) that

$$|\langle 1_L * \mu_{S_1} * \mu_{S_2} * \lambda_X^{(\ell)}, 1_{-A_2}\rangle_{L^2}| \leqslant \tfrac{1}{2}\lambda\alpha_2 b \tag{6.15}$$

for the choice of parameters $p = 2 + \log\alpha_2^{-1}$ and $\theta = \lambda^{1-1/p}/4e \asymp 1$.

Obtaining an L^2 density increment. Since the support of $\mu_{S_1} * \mu_{S_2} * \lambda_X^{(\ell)}$ is contained in $(2\ell+2)B' \subset B_{(2\ell+2)\rho}$ we have, by Lemma 3.8,

$$\langle 1_L * \mu_{S_1} * \mu_{S_2} * \lambda_X^{(\ell)}, 1_B\rangle_{L^2} = \lambda b + O(\ell\rho d b) \geqslant \tfrac{3}{4}\lambda b \tag{6.16}$$

provided that $\rho \leqslant c/(\ell d)$, which will turn out to be the case. Forming the balanced function $f_{-A_2} = 1_{-A_2} - \alpha_2 1_B$, we see from (6.15) and (6.16) that

$$|\langle 1_L * \mu_{S_1} * \mu_{S_2} * \lambda_X^{(\ell)}, f_{-A_2}\rangle_{L^2}| \geqslant \tfrac{1}{4}\lambda\alpha_2 b$$

A computation entirely analogous to that in the proof of Proposition 6.2 then shows that, choosing $\ell = C \log 2\alpha_2^{-1}$ with C large enough, we have

$$\sum_{\gamma \in \mathrm{Spec}_{1/2}(\mu_X)} |\widehat{f}_{A_2}(\gamma)|^2 \gg \alpha_2^2 b.$$

The parameters we have chosen have size $p \asymp \log 2\alpha_2^{-1}$, $\ell \asymp \log 2\alpha_2^{-1}$, and $\theta \asymp 1$. By (6.14), (6.11) and the bound $\log(2/v\widetilde{\alpha}) \ll (\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})$, we have therefore

$$\tau \geqslant \exp\Big(-C\alpha_1^{-1/2}(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^4\Big).$$

Since $\rho' \asymp v\widetilde{\alpha}/d$ and $\rho'' \asymp 1/d$, we also have $\delta''' = c\rho(v\widetilde{\alpha}/d^2)\delta$. Applying Lemma 5.2 to $A = A_2$ with $\eta = 1/2$ and some $\nu \asymp 1$, we obtain a regular Bohr set $\breve{B} \leqslant B'''$ such that

$$\begin{aligned} \|1_{A_2} * \mu_{\breve{B}}\|_\infty &\geqslant (1+c)\alpha_2, \\ \breve{d} &\leqslant d + C\alpha_1^{-1/2}(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^4, \\ \breve{\delta} &\geqslant c\rho(v\widetilde{\alpha}/d)^4\delta, \end{aligned}$$

which again is enough to conclude. □

Owing to the shape of Proposition 6.4, we now need to find arithmetic progressions in thick subsets of Bohr sets. This is precisely addressed by Sanders from [**78**, Lemma 6.7], which we now quote.

LEMMA 6.5. *Let $v \in (0,1)$ be parameter and let B be a regular Bohr set. Suppose that $v^{-1} \leqslant c\delta N^{1/d}/d$ and $A \subset B$ has relative density at least $1-v$, then A contains an arithmetic progression of length at least $4v^{-1}$.*

We now modify our iterative lemma so as to yield arithmetic progressions in upper-level sets and so as to bound the number of steps in the iteration more easily.

PROPOSITION 6.6 (Final iterative lemma). *Let $\rho, v, \omega \in (0,1)$ be parameters. Let B be a regular Bohr set and assume that $B' = B_\rho$ is regular. Suppose that*

$A_1, A_2, A_3 \subset B$ have relative densities $\alpha_1, \alpha_2, \alpha_3$, respectively, and write $\widetilde{\alpha} = \alpha_1\alpha_2\alpha_3$. Assume that $\rho \leqslant c\widetilde{\alpha}/d$,

$$v^{-1} \leqslant c\delta' N^{1/d}/d \quad \text{and} \quad 0 \leqslant \omega \leqslant \exp\Big(-C(d+\alpha_1^{-1/2})\log(2d/\rho v\widetilde{\alpha})\Big). \tag{6.17}$$

Then either

(i) *there exists a regular Bohr set $\breve{B} \leqslant B'$ such that*

$$\prod_{1\leqslant j\leqslant 3}\|1_{A_j} * \mu_{\breve{B}}\|_\infty \geqslant (1+c)\widetilde{\alpha},$$
$$\breve{d} \leqslant d + C\alpha_1^{-1/2}(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^4,$$
$$\breve{\delta} \geqslant c\rho(v\widetilde{\alpha}/d)^C\delta,$$

(ii) *or the set $\{y : 1_{A_1} * 1_{A_2} * 1_{A_3}(y) > \omega b^2\}$ contains an arithmetic progression of length at least $4v^{-1}$.*

PROOF. By Proposition 6.4 we may either find $x \in G$ such that

$$\mathcal{V} = \{y : 1_{A_1} * 1_{A_2} * 1_{A_3}(y+x) > \omega b^2\} \cap B'$$

has relative density at least $1-v$ in B', in which case we may conclude by Lemma 6.5 with $A = \mathcal{V}$; or we may obtain a regular Bohr set $\breve{B}$ such that $\|1_{A_i} * \mu_{\breve{B}}\|_\infty \geqslant (1+c)\alpha_i$ for some $i \in \{1,2\}$ and with the prescribed radius and dimension bounds. Picking j, k such that $\{i,j,k\} = \{1,2,3\}$, Lemma 6.1 then shows that

$$\prod_{1\leqslant \ell\leqslant 3}\|1_{A_\ell} * \mu_{\breve{B}}\|_\infty \geqslant (1+c)(1-O(\tfrac{\rho d}{\alpha_j}))(1-O(\tfrac{\rho d}{\alpha_k}))\widetilde{\alpha}$$

and assuming $\rho \leqslant c'\widetilde{\alpha}/d$ with c' small enough this is indeed more than $(1+c/2)\widetilde{\alpha}$. □

We are now ready for the proof of Theorem 1.6, which we quote below with adjusted notation for convenience.

PROPOSITION (Theorem 1.6). *Let $\varepsilon \in (0,1)$ be a parameter and suppose that A_1, A_2, A_3 are subsets of $\mathbb{Z}/N\mathbb{Z}$ of respective densities $\alpha_1, \alpha_2, \alpha_3$, and write $\widetilde{\alpha} =$*

$\alpha_1\alpha_2\alpha_3$. Then $A_1 + A_2 + A_3$ contains an arithmetic progression P of length at least

$$\exp\left(c\varepsilon^{1/2}\alpha_1^{1/4}(\log N)^{1/2}(\log 2\widetilde{\alpha}^{-1})^{-7/2}\right) \quad \textit{if} \quad \alpha_1(\log 2\widetilde{\alpha}^{-1})^{-14} \geqslant C(\varepsilon \log N)^{-2}$$

and such that $1_{A_1} * 1_{A_2} * 1_{A_3}(x) \geqslant N^{-\varepsilon}$ for every $x \in P$.

PROOF. The proof proceeds by iteration of Proposition 6.6. We are brief since the iteration process is very similar to that of the proof of Proposition 6.3.

We construct iteratively a sequence of regular Bohr sets $B^{(i)}$ with parameters d_i, δ_i, b_i and, for every $1 \leqslant j \leqslant 3$, a sequence of sets $A_j^{(i)} \subset B^{(i)}$ of relative density $\alpha_j^{(i)}$, and we write $\widetilde{\alpha}^{(i)} = \alpha_1^{(i)}\alpha_2^{(i)}\alpha_3^{(i)}$. We initiate the iteration with $B^{(1)} = \mathbb{Z}/N\mathbb{Z}$ and $A_j^{(1)} = A_j$ for $1 \leqslant j \leqslant 3$. At each step i we apply Proposition 6.6 to the sets $A_j^{(i)}$ with parameters ρ_i, v, ω to be determined later (note that v and ω are chosen independent of i), and in case (i) we define $B^{(i+1)} = \breve{B}^{(i)}$, while in case (ii) we stop the iteration. For every $1 \leqslant j \leqslant 3$, we pick $x_{i,j}$ so that $A_j^{(i+1)} := (A_j - x_{i,j}) \cap B^{(i+1)}$ has relative density $\alpha_j^{(i+1)} = \|1_{A_j^{(i)}} * \mu_{B^{(i+1)}}\|_\infty$ in $B^{(i+1)}$, whenever $B^{(i+1)}$ is defined.

By the density increment $\widetilde{\alpha}^{(i+1)} \geqslant (1+c)\widetilde{\alpha}^{(i)}$ from Proposition 6.6 we see that the iteration stops after at most $n = O(\log 2\widetilde{\alpha}^{-1})$ steps. We choose $\rho_i = c\widetilde{\alpha}^{(i)}/(i^2 d_i)$ such that $B_{\rho_i}^{(i)}$ is regular (via Lemma 3.7). By Lemma 6.1 we then have $\alpha_j^{(i+1)} \geqslant (1 - O(i^{-2}))\alpha_j^{(i)}$ for every i, j and therefore $\alpha_j^{(i)} \geqslant e^{-O(\sum_{i=1}^{\infty} i^{-2})}\alpha_j \gg \alpha_j$ uniformly in $1 \leqslant j \leqslant 3$ and $1 \leqslant i \leqslant n$. We then have, from the bounds of Proposition 6.6,

$$d_{i+1} \leqslant d_i + C\alpha_1^{-1/2}(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^4$$

for $i < n$ and therefore $d_i \leqslant C\alpha_1^{-1/2}(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^5$ uniformly in $1 \leqslant i \leqslant n$. Bounding crudely $\log(2/v\widetilde{\alpha}) \ll (\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})$, we also have

$$\delta_{i+1} \geqslant \exp\left(-C(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})\right)\delta_i$$

for $i < n$ and therefore $\delta_i \geqslant \exp\left(-C(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^2\right)$ uniformly in $1 \leqslant i \leqslant n$.

We now choose v and ω so that (6.17) is satisfied at every step. From the previous dimension and radius bounds, we see that a sufficient condition for v is

$$\log 2v^{-1} \leqslant \frac{c\alpha_1^{1/2}\log N}{(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^5} - C(\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})^2.$$

We choose v defined by $\log 2v^{-1} = c'\varepsilon^{1/2}\alpha_1^{1/4}(\log N)^{1/2}(\log 2\widetilde{\alpha}^{-1})^{-7/2}$ with c' small enough so as to satisfy this; since $\log 2v^{-1} \in [\log 2, +\infty)$, this requires $\alpha_1(\log 2\widetilde{\alpha}^{-1})^{-14} \geqslant C(\varepsilon \log N)^{-2}$ for a certain large enough C. Bounding again crudely $\log(2/v\widetilde{\alpha}) \ll (\log 2v^{-1})(\log 2\widetilde{\alpha}^{-1})$, we also see that a sufficient condition for ω to satisfy (6.17) is

$$\omega \leqslant \exp\Big(-C\alpha_1^{-1/2}(\log 2v^{-1})^2(\log 2\widetilde{\alpha}^{-1})^6\Big)$$

which allows for the choice $\omega = N^{-(c\varepsilon/\log 2\widetilde{\alpha}^{-1})}$ upon inserting the above expression of $\log 2v^{-1}$. From Lemma 3.4 and the choices of v and ω, we eventually obtain $\omega b_i^2 \geqslant N^{-\varepsilon}$ uniformly in $1 \leqslant i \leqslant n$. When we are in case (ii) of Proposition 6.6, we have therefore found the desired arithmetic progression. □

7. Arithmetic progressions in sumsets of sets of primes

We now consider applications of Theorems 1.5 and 1.6 to the problem of finding arithmetic progressions in $A + A + A$, for A a subset of the primes. This problem was first considered by Cui, Li and Xue in [**12**]. In that paper a connection with the original problem on arithmetic progressions in sumsets of sets of integers was outlined and exploited via the original theorem of Green on $A + A + A$, which finds an arithmetic progression of size $N^{c\alpha^2}$ in this sumset when A has density α. To obtain Theorem 1.7 we exploit the same connection, taking advantage of the slightly longer progression given by Theorem 1.5. Corollary 1.8 is obtained differently, by a direct application of Theorem 1.6.

We denote by $\log_k$ the logarithm iterated k times and we let n be a large enough integer. We also recall that when G, H are two groups, a Freiman 3-isomorphism from $A \subset G$ to $B \subset H$ is a map $\phi\colon A \to B$ such that, for every $(a_i)_{1\leqslant i\leqslant 3}$ and

$(a'_i)_{1\leqslant i\leqslant 3}$ in A^3, $\sum_i a_i = \sum_i a'_i$ if and only if $\sum_i \phi(a_i) = \sum_i \phi(a'_i)$; we refer the reader to [100, Section 5.3] for the properties of such maps. The following can be extracted from the computations of [12].

PROPOSITION 7.1. *Let $\varepsilon, \delta \in (0,1)$ and suppose that A has density α in $\{1,\dots,n\}\cap\mathcal{P}$. Then there exist an integer N such that $n/(\log n) \ll N \ll n$, a subset A' of A which is Freiman 3-isomorphic to a subset A'' of $\mathbb{Z}/N\mathbb{Z}$, a function f on $\mathbb{Z}/N\mathbb{Z}$ with support in A'', and a subset A_1 of $\mathbb{Z}/N\mathbb{Z}$ of density at least $c\alpha$ such that*

$$f*f*f(x) \geqslant \alpha^3\, 1_{A_1}*1_{A_1}*1_{A_1}(x) - O(\varepsilon+\delta^{1/2}) \qquad (x\in G) \tag{7.1}$$

provided $C(\log_4 N)/(\log_2 N) \leqslant (\varepsilon/2\pi)^{C\delta^{-5/2}}$.

PROPOSITION 7.2. *Let $\varepsilon, \delta \in (0,1)$ and suppose that A has density α in $\{1,\dots,n\}\cap\mathcal{P}$. Then there exist an integer N such that $n^{1/2} \ll N \ll n$, a subset A' of A which is Freiman 3-isomorphic to a subset A'' of $\mathbb{Z}/N\mathbb{Z}$, a function g on $\mathbb{Z}/N\mathbb{Z}$ with support in A'' and a subset A_1 of $\mathbb{Z}/N\mathbb{Z}$ of density at least $c\alpha^2$ such that*

$$g*g*g(x) \geqslant \alpha^3\, 1_{A_1}*1_{A_1}*1_{A_1}(x) - O(\varepsilon+\delta^{1/2}) \qquad (x\in G)$$

provided $\delta^{-5/2}\log 2\varepsilon^{-1} \leqslant c\log N$.

PROOF OF THEOREM 1.7. To obtain the first estimate we apply Proposition 7.1. Since A_1 has density at least $c\alpha$, we know by Theorem 1.5 that $A_1+A_1+A_1$ contains an arithmetic progression P of length at least $N^{c\alpha/(\log 2\alpha^{-1})^5}$ such that, for every $x\in P$,

$$1_{A_1}*1_{A_1}*1_{A_1}(x) \geqslant \exp\Big(-C\alpha^{-1}(\log 2\alpha^{-1})^7\Big).$$

Choosing $\varepsilon=\delta=\exp(-C'\alpha^{-1}(\log 2\alpha^{-1})^7)$ with C' large enough it then follows from (7.1) that $f*f*f(x)>0$ for all $x\in P$, and therefore that $P\subset A''+A''+A''$. Pulling

back to $A' \subset A$ by the Freiman isomorphism we are done provided $\delta^{-5/2} \log 2\varepsilon^{-1} \leqslant c \log_3 N$, which is satisfied for $\alpha \geqslant C(\log_5 N)^7/\log_4 N$.

To obtain the second estimate we apply Proposition 7.2, where this time A_1 has density at least $c\alpha^2$. Theorem 1.5 then yields a progression $P \subset A_1 + A_1 + A_1$ of length at least $N^{c\alpha^2/(\log 2\alpha^{-1})^5}$ such that

$$1_{A_1} * 1_{A_1} * 1_{A_1}(x) \geqslant \exp\Big(-C\alpha^{-2}(\log 2\alpha^{-1})^7\Big),$$

and choosing $\delta = \varepsilon = \exp(-C'\alpha^{-2}(\log 2\alpha^{-1})^7)$ we may conclude as before provided

$$e^{C\alpha^{-2}(\log 2\alpha^{-1})^7} \leqslant c \log N.$$

This is certainly satisfied for $\alpha \geqslant C(\log_3 N)^{7/2}/(\log_2 N)^{1/2}$. □

Proof of Corollary 1.8. The projection $\pi\colon \mathbb{Z} \to \mathbb{Z}/6N\mathbb{Z}$ is a Freiman 3-isomorphism from $A \subset \{1, \dots, N\}$ to $A' \coloneqq \pi(A)$ which preserves arithmetic progressions. Note that A' has density $\gg \alpha/\log N$ in $\mathbb{Z}/6N\mathbb{Z}$. Applying Theorem 1.6 with $A = B = C = A'$, $\varepsilon = \frac{1}{2}$ and pulling back to $\mathbb{Z}$ then concludes the proof. □

8. Remarks and conclusion

There is a strong parallel between the quantitative results one can obtain about arithmetic progressions in sumsets and on Roth's theorem by the density-increment strategy of [81]. Indeed the limitation in the range of density in both problems is similar. To see this, consider a subset A of $\mathbb{Z}/N\mathbb{Z}$ of density α. Sanders [81] then showed that when $\alpha \geqslant (\log N)^{-1+o(1)}$, there exists a nontrivial three-term arithmetic progression in A, which Bloom [2] generalized to show (in particular) that for $\alpha \geqslant (\log N)^{-2+o(1)}$, any translation-invariant equation in four variables has a nontrivial solution in A. By comparison, the same density-increment strategy applied to our problem can be made to obtain a long progression in $A + A$ in the range $\alpha \geqslant (\log N)^{-1+o(1)}$ (although this is not pursued here, since the argument of [9] is simpler in this case) and, by Theorem 1.6, it yields one in $A + A + A$ for

$\alpha \geqslant (\log N)^{-2+o(1)}$. It is therefore likely that any improvement of this technique would result in a better density dependency in both problems.

Chapitre IV. Arithmetic progressions in sets of small doubling

---

Author: Kevin Henriot.

Abstract: We show that if a finite, large enough subset A of an arbitrary abelian group satisfies the small doubling condition $|A+A| \leqslant (\log|A|)^{1-\varepsilon}|A|$, then A must contain a three-term arithmetic progression whose terms are not all equal, and $A+A$ must contain an arithmetic progression or a coset of a subgroup, either of which of size at least $\exp\left[c(\log|A|)^{\delta}\right]$. This extends analogous results obtained by Sanders and, respectively, by Croot, Laba and Sisask in the case where the group is $\mathbb{Z}^s$ or $\mathbb{F}_q^n$.

1. Introduction

Our aim in this work is to generalize two types of results of additive combinatorics usually stated for dense subsets of the integers, namely Roth's theorem [69] and Bourgain's theorem on long arithmetic progressions in sumsets [4], to the case where the sets only have small doubling and live in an arbitrary abelian group. As in previous work of this nature [75, 80, 93, 94], our motivation is to provide a link between two types of additive structure: small doubling on the one hand, and containment of arithmetic progressions in the set or its sumset on the other hand. Since the result we seek is known qualitatively by the modelling methods of Green and Ruzsa [32], we focus on the quantitative bounds that may be obtained for it.

Concerning the first topic of Roth's theorem, we start by recalling the state-of-the-art bounds, which we state in the setting of a cyclic group. Here a k-term arithmetic progression in an abelian group is defined as a tuple $(x_1, \dots, x_k)$, where $x_1, \dots, x_k$ are group elements such that $x_2 - x_1 = \cdots = x_k - x_{k-1}$, and we say that it is trivial when $x_1, \dots, x_k$ are all equal, and proper when they are all distinct; note that when the group has odd order every nontrivial three-term arithmetic progression is proper. The breakthrough work of Sanders [81] then, building on earlier work of Bourgain [5], has established that given a large enough, odd integer N, every subset of $\mathbb{Z}/N\mathbb{Z}$ of density at least $(\log N)^{-1+o(1)}$ contains a proper three-term arithmetic progression. Under a density hypothesis, the generalization to finite abelian groups is not very challenging: indeed it can be essentially read out of [81] that any set of density at least $(\log |G|)^{-1+o(1)}$ in a finite abelian group G of odd order contains a proper three-term arithmetic progression.

However, the situation is more complex when we only assume that the set in question, say A, has small doubling in the sense that $|A + A| \leqslant K|A|$. Since subsets of density α of a finite abelian group have doubling at most $K = \alpha^{-1}$, this includes the previous situation. We would then like to show that $K \leqslant (\log |A|)^{1-o(1)}$ forces A to contain a proper three-term arithmetic progression, which would truely generalize the dense case, however this is not not obvious even in the case where A is a set of integers. Indeed the direct approach, which proceeds by combining the standard Ruzsa modelling lemma [75] with the bounds for Roth's theorem from [81], only yields an admissible range of $K \leqslant (\log |A|)^{1/4-o(1)}$. This is precisely what led Sanders [80] to design a more subtle approach which, for sets of integers, yields the range we seek.

THEOREM 1.1 (Sanders). *There exists an absolute constant $c > 0$ such that the following holds. Suppose that A is a finite set of integers such that*[1]

$$|A + A| \leqslant c(\log |A|)(\log\log |A|)^{-8} \cdot |A|.$$

Then A contains a proper three-term arithmetic progression.

This does not appear explicitely in the literature, but follows more or less directly from inserting Ruzsa's modelling bound [75] into the argument of [80], taking also into account the latest bounds for Roth's theorem [81]; we describe this in more detail at the end of the article. By this procedure, one can actually obtain a version of Theorem 1.1 for any group with good modelling in the sense of [32]. In the general abelian case, where available modelling arguments are by necessity much weaker [32], Sanders [80] also improves substantially on the bounds that would follow from a direct modelling approach.

THEOREM 1.2 (Sanders). *There exists an absolute constant $c > 0$ such that the following holds. Suppose that A is a finite subset of an abelian group such that*

$$|A + A| \leqslant c(\log |A|)^{1/3}(\log\log |A|)^{-1} \cdot |A|.$$

Then A contains a nontrivial three-term arithmetic progression.

Note that the conclusion changed to yield a nontrivial arithmetic progression only; we say more on this later. The loss in the exponent of $\log |A|$ in comparison with the previous case is due to a limitation of the results on modelling; indeed via [32] it is only possible to Freiman-embed a set A of doubling K into a finite abelian group where its image has density $\exp[-CK^2 \log K]$. A construction by Green and Ruzsa [32] further shows that any modelling result of this type will feature an exponential loss in $\sqrt{K}$, at least if we insist on embedding the whole set.

[1] Throughout this introduction, we make the tacit assumption that all quantities appearing inside a double logarithm are at least e^e in size.

Fortunately, in a recent major advance on the polynomial Freiman-Ruzsa conjecture, Sanders [83] managed to sidestep this issue and obtained a correlation result which may be viewed as another form of modelling. This result may be applied to our situation to recover a range of doubling matching the current bounds for Roth's theorem, for arbitrary abelian groups; this is the first observation of this paper.

THEOREM 1.3. *There exists an absolute constant $c > 0$ such that the following holds. Suppose that A is a finite subset of an abelian group such that*

$$|A + A| \leqslant c(\log |A|)(\log \log |A|)^{-7} \cdot |A|.$$

Then A contains a nontrivial three-term arithmetic progression.

Here we say more on the issue of 2-torsion, which was already discussed by Sanders in [80]. In general, a set A contains a nontrivial degenerate arithmetic progression (x, y, x) if and only if $A - A$ contains an element of order 2; therefore in that case, Theorems 1.2 and 1.3 give only trivial information. Obtaining proper progressions in every case where it is possible (this excludes groups such as $\mathbb{F}_2^n$) is a thorny issue that has only been successfully adressed in work of Lev [60] and Sanders [79] in cases where the group rank is not too large; here we do not consider this issue.

The second topic we consider is that of long arithmetic progressions in sumsets, initiated by Bourgain [4] and further developed by Green [29]. Basing themselves on a fundamental new technique introduced by Croot and Sisask [11], these two last authors together with Laba [9] obtained a remarkable extension of Green's result, which furthermore already works under a small doubling hypothesis.

THEOREM 1.4 (Croot, Laba, Sisask). *There exists an absolute constant $c > 0$ such that the following holds. Let $K, L \geqslant 1$ be parameters, and suppose that A, B are finite sets of integers such that $|A + B| \leqslant K|A|$ and $|A + B| \leqslant L|B|$. Then $A + B$*

contains an arithmetic progression of length at least

$$\exp\left[c\left(\frac{\log|A+B|}{K(\log L)^3}\right)^{1/2}\right] \quad \textit{provided} \quad K\log^5(L\log|A|)\leqslant c\log|A+B|.$$

From the methods of [9], one can easily deduce that an analog result holds for subsets A and B of density α and β of a finite abelian group, with α^{-1} and β^{-1} in place of K and L. Therefore we focus again on the case of small doubling in an arbitrary abelian group, to which the argument of [9] does not extend as it relies on a two-sets version of Ruzsa modelling [75]. The coveted generalization of Theorem 1.4 may however be recovered, again by using the Bogolyubov-Ruzsa lemma from [83], and establishing this is the second aim of this paper. Note that in the general abelian setting, we need to adapt the type of structure sought to allow for both cosets of subgroups and arithmetic progressions.

THEOREM 1.5. *There exists an absolute constant $c>0$ such that the following holds. Let $K\geqslant 1$ be a parameter and suppose that A is a finite subset of an abelian group such that $|A+A|\leqslant K|A|$. Then $A+A$ contains a set, which is either a proper arithmetic progression or a coset of a subgroup, of size at least*

$$\exp\left[c\left(\frac{\log|A|}{K(\log K)^3}\right)^{1/2}\right] \quad \textit{provided} \quad K\leqslant\frac{c\log|A|}{(\log\log|A|)^5}.$$

This recovers Theorem 1.4 in the symmetric case $A=B$, since in $\mathbb{Z}$ every nontrivial subgroup is infinite. We restrict to the symmetric case for simplicity; it seems feasible to obtain an asymmetric result of the shape of Theorem 1.4 from the methods of this paper, however we do not pursue this here.

Finally, we mention an application of results on arithmetic progressions in sets of small doubling, to the asymptotic size of restricted sumsets. This application was first observed independently by Schoen [86] and Hegyvári et al. [49] in the setting of integers, and later quantitatively strengthened by Sanders [80] in the more general setting of abelian groups. We write $A\mathbin{\widehat{+}}A$ for the set of sums of distinct elements of A below.

COROLLARY 1.6. *Suppose that A is a finite nonempty subset of an abelian group. Then*

$$|A \mathbin{\widehat{+}} A| \geqslant \big(1 - (\log |A|)^{-1+o(1)}\big)|A + A|.$$

This improves upon the exponent $-\frac{1}{3}$ on the logarithm obtained by Sanders [**80**] via Theorem 1.2, since Theorem 1.3 is used instead. Note that by Behrend's construction [**64**], the restricted sumset may have size as low as $(1 - e^{-c\sqrt{\log |A|}})|A + A|$ and therefore the bounds for this problem match those for Roth's theorem closely.

Finally, we remark that by the finite modelling argument of Green and Ruzsa [**32**, Lemma 2.1], it suffices to prove all our results in the case where the group is finite abelian, and therefore we work under that hypothesis for the rest of the paper. This concludes our introduction and we discuss the structure of this paper in the next section.

Funding. This research was supported by a *contrat doctoral* from Université Paris 7 and by the ANR Caesar ANR-12-BS01-0011.

2. Overview

In this section we sketch the argument behind our results and outline the structure of this paper. We use the symbols $\approx$ and $\gtrsim$ to indicate statements that hold true up to certain negligible factors.

The first logical step in the proof of Theorem 1.3 consists in applying the correlation version of Sanders' Bogolyubov-Ruzsa lemma [**83**] (Proposition 7.1) to deduce that a set A of doubling K has density $\asymp 1/K$ in (a translate of) a large Bourgain system B, a group-like object whose properties are recalled in Section 4. The second step is to obtain an efficient local version of Roth's theorem (Proposition 6.1), which, roughly saying, asserts that a set A of density $\alpha \gtrsim (\log |B|)^{-1}$ in a large Bourgain system B contains many arithmetic progressions, and therefore a nontrivial one. This may be applied to the previous system B, for

which $|B| \approx |A|$ and $\alpha \asymp 1/K$, under the condition $K \lesssim \log |A|$, thereby establishing Theorem 1.3. The local Roth theorem is developed in Section 6, drawing on analytic tools from Section 5, and it is combined in the preceding fashion with the correlation Bogolyubov-Ruzsa lemma in Section 7.

To derive Theorem 1.5, we need to obtain instead a local version of an almost-periodicity lemma of Croot et al. [**9**] (Proposition 8.4), drawing again on the tools of Section 5. This process, carried out in Section 8, requires a somewhat simpler version of Sanders' Bogolyubov-Ruzsa lemma (Proposition 8.1) which deduces containment of a large Bourgain system in the sumset $2A - 2A$ from the hypothesis that A has small doubling, and the rest of the argument follows the strategy of [**9**].

Finally, to illustrate some of the above ideas, we showcase the proof of Theorem 1.3 in the model setting of $\mathbb{F}_3^n$, where the proof of Sanders' Bogolyubov-Ruzsa lemma [**83**] simplifies substantially. As an added benefit, the formidable bounds of Bateman and Katz [**1**] for caps in $\mathbb{F}_3^n$ yield a larger admissible range of doubling in this setting. The notation used in the proof is introduced in Section 3.

THEOREM 2.1. *There exist positive absolute constants c and ε such that the following holds. Suppose that A is a subset of $\mathbb{F}_3^n$ such that*

$$|A + A| \leqslant c(\log |A|)^{1+\varepsilon} \cdot |A|.$$

Then A contains a proper three-term arithmetic progression.

PROOF. Write $K = |A+A|/|A|$, so that we are assuming that $K \leqslant c(\log |A|)^{1+\varepsilon}$. The proof of [**32**, Proposition 6.1] readily adapts to $\mathbb{F}_3^n$, and shows that A is Freiman-isomorphic to a subset of doubling K and density at least K^{-4} of another finite field $\mathbb{F}_3^m$, which we identify with A from now on. By examining the proof of [**83**, Theorem A.1], which works equally well in $\mathbb{F}_3^m$, one may deduce that there exist a measure μ and a subspace V of $\mathbb{F}_3^m$ of codimension at most $C(\log K)^4$ such

that

$$\langle 1_A * \mu_V * \mu_{A+A} * \mu, \mu_A \rangle_{L^2} \geqslant \tfrac{1}{2}\mu_G(A)/\mu_G(A+A).$$

By the definition of K, and upon applying Hölder's and Young's inequalities, we obtain

$$\begin{aligned} \tfrac{1}{2K} &\leqslant \langle 1_A * \mu_V * \mu_{A+A} * \mu, \mu_A \rangle_{L^2} \\ &\leqslant \|1_A * \mu_V * \mu_{A+A} * \mu\|_\infty \|\mu_A\|_{L^1} \\ &\leqslant \|1_A * \mu_V\|_\infty. \end{aligned}$$

Therefore we may find x such that $A' = (A-x) \cap V$ has density at least $\frac{1}{2K}$ in V. Since V has codimension at most $C(\log K)^4$, it has size at least $|G|^{1/2}$ in our range of K. Applying [1, Theorem 1.1] to A', we are then ensured to find a proper three-term arithmetic progression in A' provided

$$\tfrac{1}{2K} \geqslant C(\log |V|)^{-(1+\varepsilon)}$$

and this concludes the proof since $\log |V| \asymp \log |A|$. □

3. Notation

In this section we introduce the notation used throughout the article.

Ambient group. We let G denote a fixed, finite abelian group. The arguments of later sections all take place in this group unless otherwise stated.

$\mathbb{Z}$-actions. The group G is naturally equipped with a structure of $\mathbb{Z}$-module, and we let $k \cdot x$ denote the action of a scalar $k \in \mathbb{Z}$ on an element $x \in G$. For a subset X of G and a subset I of $\mathbb{Z}$, we further write

$$k \cdot X = \{k \cdot x : x \in X\} \quad \text{and} \quad I \cdot x = \{k \cdot x : k \in I\}.$$

Note that $\cdot$ is also used in other places for the regular multiplication of complex numbers, however it should be clear from the context which one is meant.

Functions. We define the averaging operator over a subset X of G, which acts on the space of functions $f : G \to \mathbb{C}$, by $\mathbb{E}_X f = |X|^{-1} \sum_{x \in X} f(x)$, and we write $\mathbb{E}_{x \in X} f(x)$ when we want to keep the variable explicit. It is also convenient to introduce the operator of translation on a function f defined by $\tau_x f(u) = f(x+u)$ for all $x, u \in G$. We furthermore define the support of f as $\mathrm{Supp}(f) = \{x \in G : f(x) \neq 0\}$. On the physical space, we use the normalized counting measure so that for functions $f, g : G \to \mathbb{C}$, we let

$$\begin{aligned}
&(L^p\text{-norm}) && \|f\|_{L^p} = (\mathbb{E}_G |f|^p)^{1/p}, \\
&(\text{Scalar product}) && \langle f, g \rangle_{L^2} = \mathbb{E}_G f\bar{g}, \\
&(\text{Convolution}) && f * g(x) = \mathbb{E}_{y \in G} f(y) g(x-y) \qquad \forall x \in G.
\end{aligned}$$

We occasionally write $\|f\|_p$ for $\|f\|_{L^p}$, and we let $f^{(\ell)}$ denote the convolution of f with itself ℓ times.

Measures. We identify measures μ on G with functions $\mu : G \to \mathbb{R}_+$ via the identity $\mu(\{x\}) = |G|^{-1}\mu(x)$, so that $\mu(E) = \langle 1_E, \mu \rangle_{L^2}$ for every subset E of G. We only consider probability measures; in other words, we always assume that $\|\mu\|_{L^1} = 1$. We write μ_A for the measure defined by $\mu_A(E) = |E \cap A|/|A|$ for every set E, which under our identification corresponds to the function $\mu_A = \mu_G(A)^{-1} 1_A$.

Fourier transform. The Fourier transform over finite abelian groups is now a standard tool of additive combinatorics. It is very well explained for example in [45], and here we only recall its main properties.

Write $\mathbb{U}$ for the unit circle, then the dual group $\widehat{G}$ is defined as the set of morphisms from G to $\mathbb{U}$, called characters, and the Fourier transform of a function $f : G \to \mathbb{C}$ is defined by $\widehat{f}(\gamma) = \langle f, \gamma \rangle_{L^2}$ at every character γ. We write $(f)^\wedge$ for the Fourier transform of f when f has a complicated expression.

We define the summation operator over a subset Δ of $\widehat{G}$, which acts on the space of functions $F : \widehat{G} \to \mathbb{C}$, by $\sum_\Delta F = \sum_{\gamma \in \Delta} F(\gamma)$. On the Fourier space, we

use the counting measure so that for functions $F, G : \widehat{G} \to \mathbb{C}$, we let

$$\begin{aligned} &(\ell^p\text{-norm}) && \|F\|_{\ell^p} = \left(\sum_{\widehat{G}} |F|^p\right)^{1/p}, \\ &(\text{Scalar product}) && \langle F, G\rangle_{\ell^2} = \sum_{\widehat{G}} F\,\overline{G}. \end{aligned}$$

The three classic formulæ of harmonic analysis then read as follows:

$$\begin{aligned} &(\text{Fourier inversion}) && f = \sum_{\widehat{G}} \widehat{f}(\gamma)\gamma, \\ &(\text{Parseval formula}) && \langle f, g\rangle_{L^2} = \langle \widehat{f}, \widehat{g}\,\rangle_{\ell^2}, \\ &(\text{Convolution identity}) && (f * g)^\wedge = \widehat{f}\cdot\widehat{g}\,. \end{aligned}$$

Other. We let c and C denote absolute positive constants, which may take different values at each occurence. Given nonnegative functions f and g, we let $f = O(g)$ or $f \ll g$ indicate the fact that there exists a constant C such that $f \leqslant Cg$, and we let $f = \Theta(g)$ or $f \asymp g$ indicate that $f \ll g$ and $g \ll f$ hold simultaneously. We also write $\ell(x) = \log(e/x)$ for $x \geqslant 1$, since this quantity arises often in our computations. Note finally that in many occurences of logarithms throughout the paper, one should replace $\log x$ by $\log ex$ for the results to be formally correct in all ranges of parameters; we leave this as a mental task to the reader to alleviate the notation. Other notation in this paper is introduced in the relevant section as needed.

4. Bourgain systems

In this section we recall the theory of Bourgain systems, which was introduced by Green and Sanders [33] as a generalization of the Bohr set technology of Bourgain [5]. In a sense these systems are the most general class of sets for which the strategy of density increment on Bohr sets, pioneered by Bourgain [5], may be carried out. What is needed for such an undertaking is for the set to behave approximately

like a d-dimensional ball with respect to dilation, as axiomatized in the following definition.

DEFINITION 4.1 (Bourgain system). *A Bourgain system of dimension d is a family of sets $\mathcal{B} = (B_\rho)_{\rho>0}$, where B_ρ are subsets of G such that, for all positive ρ and ρ',*

$$\begin{array}{lll} \textit{(containment of } 0\textit{)} & 0 \in B_\rho & \\ \textit{(symmetry)} & -B_\rho = B_\rho & \\ \textit{(nesting)} & B_\rho \subset B_{\rho'} & \textit{if} \quad \rho \leqslant \rho' \\ \textit{(additive closure)} & B_\rho + B_{\rho'} \subset B_{\rho+\rho'} & \\ \textit{(}2^d\textit{-covering)} & \exists X_\rho : B_{2\rho} \subset X_\rho + B_\rho & \textit{and} \quad |X_\rho| \leqslant 2^d. \end{array}$$

We write $B = B_1$, and we define the density of $\mathcal{B}$ as $b = |B|/|G|$.

We let the sets B_ρ, and sometimes also the dimension d and the density b, be defined implicitely whenever we introduce a Bourgain system $\mathcal{B}$. We now describe two important classes of Bourgain systems: Bohr sets and coset progressions. To define the former, we consider the multiplicative analog $\|\cdot\|_{\mathbb{U}}$ on the unit circle of the usual pseudo-norm $\|\cdot\|_{\mathbb{T}} = d(\cdot, \mathbb{Z})$ on the torus, defined by $\|e(\theta)\|_{\mathbb{U}} = \|\theta\|_{\mathbb{T}}$ for every $\theta \in \mathbb{T}$.

DEFINITION 4.2 (Bohr set). *Suppose that $\Gamma \subset \widehat{G}$ and $\delta > 0$. The Bohr set of frequency set Γ and radius δ is*

$$B = B(\Gamma, \delta) = \{x \in G : \|\gamma(x)\|_{\mathbb{U}} \leqslant \delta\}.$$

The dimension of B is $d = |\Gamma|$. We define the dilate of B by $\rho > 0$ as the set $B_\rho = B(\Gamma, \rho\delta)$, and the Bohr system induced by B as the system $\mathcal{B} = (B_\rho)_{\rho>0}$.

The usual bounds for the size and growth of a Bohr set allow us to quickly estimate the dimension and density of the Bourgain system it induces.

LEMMA 4.3. *The system $\mathcal{B}$ induced by a Bohr set of dimension d and radius $\delta \leqslant 1$ is a Bourgain system $\mathcal{B}$ of dimension at most $6d$ and density at least δ^d.*

PROOF. The first four properties of a Bourgain system are easy to check. Further, by three applications of [100, Lemma 4.20] we obtain $|B_{4\rho}| \leqslant 2^{6d}|B_{\rho/2}|$, and therefore by Ruzsa's covering lemma we may find a set X_ρ such that

$$B_{2\rho} \subset X_\rho + B_{\rho/2} - B_{\rho/2} \subset X_\rho + B_\rho$$

and $|X_\rho| \leqslant |B_{2\rho} + B_{\rho/2}|/|B_{\rho/2}| \leqslant 2^{6d}$. Working through the argument in that reference, one could extract a better bound 2^{2d}, but this would not affect our end results much. The bound on the density may be read directly from [100, Lemma 4.20]. An alternate reference for these estimates is [52, Section 5]. □

In our definition of a coset progression, we write $[x, y]_{\mathbb{Z}} = \{n \in \mathbb{Z} : x \leqslant n \leqslant y\}$ for reals $x \leqslant y$.

DEFINITION 4.4 (Coset progression). *Let $L \in \mathbb{R}_+^d$ and $\omega \in G^d$ where $d \geqslant 1$, and let H be a subgroup of G. The coset progression of dimension d determined by L, ω, H is*

$$M = M(L, \omega, H) = [-L_1, L_1]_{\mathbb{Z}} \cdot \omega_1 + \cdots + [-L_d, L_d]_{\mathbb{Z}} \cdot \omega_d + H.$$

We define the dilate of M by $\rho > 0$ as $M_\rho = M(\rho L, \omega, H)$, and the coset progression system induced by M as the system $\mathcal{M} = (M_\rho)_{\rho>0}$.

The dimension of the Bourgain system induced by a coset progression may be estimated by a simple covering argument.

LEMMA 4.5. *The system $\mathcal{M}$ induced by a d-dimensional coset progression M is a Bourgain system of dimension at most $3d$.*

PROOF. It is again rather simple to derive the first four properties of a Bourgain system for $\mathcal{M}$, and we now concern ourselves with the fifth. The dilate of M by

$\rho > 0$ is

$$M_\rho = [-\rho L_1, \rho L_1]_{\mathbb{Z}} \cdot \omega_1 + \cdots + [-\rho L_d, \rho L_d]_{\mathbb{Z}} \cdot \omega_d + H.$$

To obtain the covering property, first observe that for any $k \in \mathbb{N}_{\geqslant 0}$, one may cover the interval $[-k, k]_{\mathbb{Z}}$ by three translates of $[-\frac{k}{2}, \frac{k}{2}]_{\mathbb{Z}}$ (this is sharp for k odd), and that this still holds for any real $k \geqslant 0$. Therefore, for every $1 \leqslant i \leqslant d$, we may find a set T_i with $|T_i| \leqslant 3$ such that $[-2\rho L_i, 2\rho L_i]_{\mathbb{Z}} \subset T_i + [-\rho L_i, \rho L_i]_{\mathbb{Z}}$. Consequently, for any $\rho > 0$ we have a covering

$$M_{2\rho} \subset \bigcup_{t \in T_1 \times \cdots \times T_d} (t_1 \cdot \omega_1 + \cdots + t_d \cdot \omega_d + M_\rho) = X_\rho + M_\rho$$

for a certain set X_ρ of size at most $|T_1| \cdots |T_d| \leqslant 3^d$. □

With these examples covered, we now work exclusively within the framework of Bourgain systems. We start by defining a few basic operations on these systems.

LEMMA 4.6 (Dilation). *Suppose that $\lambda \in (0, 1]$ and that $\mathcal{B}$ is a Bourgain system of dimension d and density b. Then the dilated system $\mathcal{B}_\lambda = \left(B_{\lambda\rho}\right)_{\rho>0}$ is a Bourgain system of dimension at most d and density at least $(\lambda/2)^d \cdot b$.*

PROOF. Let $\lambda \in (0, 1]$, and choose $k \geqslant 0$ such that $2^{-(k+1)} < \lambda \leqslant 2^{-k}$. By the covering property of Definition 4.1, we have $|B_\rho| \leqslant 2^d |B_{\rho/2}|$ for every $\rho > 0$, from which it follows by iteration that $|B| \leqslant 2^{(k+1)d} |B_{1/2^{k+1}}| \leqslant (2/\lambda)^d |B_\lambda|$. That $\mathcal{B}_\lambda$ is a d-dimensional Bourgain system is obvious, and the bound on the density follows from the previous computation. □

DEFINITION 4.7 (Sub-Bourgain systems). *Suppose that $\mathcal{B}$ and $\mathcal{B}'$ are two Bourgain systems. We say that $\mathcal{B}$ is a sub-Bourgain system of $\mathcal{B}'$, and we write $\mathcal{B} \leqslant \mathcal{B}'$, when $B_\rho \subset B'_\rho$ for all $\rho > 0$. For $\lambda \in (0, 1]$, we also write $\mathcal{B} \leqslant_\lambda \mathcal{B}'$ when $\mathcal{B} \leqslant \mathcal{B}'_\lambda$.*

The properties of an intersection of Bourgain systems were derived in [**80**, Lemma 3.4], whose proof we reproduce here for completeness.

LEMMA 4.8 (Intersection). *Suppose that $\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(k)}$ are Bourgain systems of dimensions $d_1, \dots, d_k$ and densities $b_1, \dots, b_k$. Then the intersection system*

$$\mathcal{B}_1 \wedge \cdots \wedge \mathcal{B}_k = (B_\rho^{(1)} \cap \cdots \cap B_\rho^{(k)})_{\rho>0}$$

is a Bourgain system of dimension at most $2(d_1 + \cdots + d_k)$ and of density at least $4^{-(d_1+\cdots+d_k)} b_1 \cdots b_k$.

PROOF. The first four properties of a Bourgain system are again easy to check, and we now consider the covering property. Let $\rho > 0$. For each $1 \leqslant i \leqslant k$, apply the covering property of $\mathcal{B}^{(i)}$ twice to obtain a set T_i of size at most 4^{d_i} such that $B_{2\rho}^{(i)} \subset T_i + B_{\rho/2}^{(i)}$. Distributing intersection over union, we have then

$$\bigcap_{1\leqslant i\leqslant d} B_{2\rho}^{(i)} = \bigcup_{(t_1,\dots,t_k)\in T_1\times\cdots\times T_k} \bigcap_{1\leqslant i\leqslant k} \left(t_i + B_{\rho/2}^{(i)}\right).$$

Now pick an element $x(t)$ in each nonempty intersection $\bigcap_i (t_i + B_{\rho/2}^{(i)})$. Then for each element x of $\bigcap_i B_{2\rho}^{(i)}$, we may find an element $t \in \prod_i T_i$ such that

$$x - x(t) \in \bigcap_i \left(B_{\rho/2}^{(i)} - B_{\rho/2}^{(i)}\right) \subset \bigcap_i B_\rho^{(i)}.$$

This yields the desired covering with X_ρ defined as the set of all $x(t)$.

To estimate the density of the intersection, first apply Ruzsa's covering lemma for each $1 \leqslant i \leqslant k$ to obtain a covering of the form

$$G \subset T_i + B_{1/4}^{(i)} - B_{1/4}^{(i)} \subset T_i + B_{1/2}^{(i)}$$

where T_i is a set of size $|T_i| \leqslant 4^{d_i} b_i^{-1}$. From $G \subset \cap_i (T_i + B_{1/2}^{(i)})$, it follows that

$$G = \bigcup_{(t_1,\dots,t_k)\in T_1\times\cdots\times T_k} \bigcap_{1\leqslant i\leqslant k} \left(t_i + B_{1/2}^{(i)}\right) = \bigcup_{t\in T_1\times\cdots\times T_k} A(t)$$

where $A(t)$ are sets satisfiying $A(t) - A(t) \subset \bigcap_i B^{(i)}$. By the pigeonhole principle, we may also find a point t such that

$$|A(t)| \geqslant \frac{|G|}{|T_1| \cdots |T_k|} \geqslant 4^{-(d_1+\cdots+d_k)} b_1 \cdots b_k |G|,$$

which yields the desired density estimate since $|A(t) - A(t)| \geqslant |A(t)|$. □

We consider one last operation on Bourgain systems; since it is so simple we leave it as an exercise to the reader.

LEMMA 4.9 (Homomorphic image). *Suppose that $\mathcal{B}$ is a Bourgain system of dimension d, and ϕ is an endomorphism of G. Then the image system $\phi(\mathcal{B}) = \big(\phi(B_\rho)\big)_{\rho>0}$ is a Bourgain system of dimension at most d.*

Finally, we recall the essential notion of regularity introduced by Bourgain [5] for Bohr sets, and which has a natural analogue for Bourgain systems. We let[2] $C_0 = 2^5$ and $C_1 = 2^6$ in what follows for definiteness, although the exact values are unimportant.

DEFINITION 4.10 (Regular Bourgain system). *We say that a Bourgain system $\mathcal{B}$ of dimension d is regular when, for every $|\rho| \leqslant \frac{1}{C_0 d}$,*

$$1 - C_0|\rho|d \leqslant \frac{|B_{1+\rho}|}{|B|} \leqslant 1 + C_0|\rho|d.$$

In practice one can always afford to work with regular Bourgain systems, as is the case with Bohr sets, via [80, Proposition 3.5] which we now quote.

LEMMA 4.11. *Suppose that $\mathcal{B}$ is a Bourgain system. Then there exists $\lambda \in \left[\frac{1}{2}, 1\right]$ such that $\mathcal{B}_\lambda$ is regular.*

The regularity computations in subsequent sections rely on the following L^1 estimate.

[2] These precise constants, featured in subsequent lemmas, are derived in [52, Section 6].

LEMMA 4.12. *Suppose that $\mathcal{B}$ is a regular Bourgain system of dimension d and μ is a measure on G with support in B_ρ, where $0 < \rho \leqslant \frac{1}{C_1 d}$. Then*

$$\|\mu_B * \mu - \mu_B\|_{L^1} \leqslant C_1 \rho d.$$

PROOF. For every $y \in B_\rho$, the function $\mu_{y+B} - \mu_B$ has support in $B_{1+\rho} \smallsetminus B_{1-\rho}$, so that

$$\|\mu_{y+B} - \mu_B\|_{L^1} \leqslant \frac{|B_{1+\rho}| - |B_{1-\rho}|}{|B|} \leqslant 2C_0 \rho d.$$

Averaging over $y \in G$ with weights $\mu(y)$, and using the triangle inequality, we recover the desired estimate. □

5. Spectral analysis on Bourgain systems

This section is concerned with collecting all the analytic information we need about the large spectrum of the indicator functions of certain sets. The main task is to obtain a large structured set on which all characters of the large spectrum take values close to 1, since such a set may be later used for purposes of a density-increment-based iteration, or to locate long arithmetic progressions.

When considering indicator functions of subsets of Bohr sets, the information we seek is provided by the spectral analysis developed by Sanders [82], and the aim of this section is therefore to obtain a similar analysis for Bourgain systems. Note that such a process was already carried out in the earlier article [80], however we benefit here from the more efficient analysis of the local spectrum from [82]. To be specific, there is now a local analog of Chang's bound [82, Lemma 4.6] which supersedes the earlier local analog of Bessel's inequality [80, Proposition 4.4]. We now give the precise statements, and in that regard it is useful to recall the following definitions.

DEFINITION 5.1 (Annihilation). *Let $\nu \in (0, 2]$ be a parameter, and suppose that T is a subset of G and Δ is a subset of $\widehat{G}$. We say that Δ is ν-annihilated by T*

when

$$|1 - \gamma(t)| \leqslant \nu \quad \textit{for all } t \in T \textit{ and } \gamma \in \Delta.$$

When $\mathcal{B}$ is a Bourgain system, we say that it ν-annihilates Δ when B does.

The quantity we seek to annihilate is then the following.

DEFINITION 5.2 (Large spectrum). *Suppose that $\eta \in (0,1]$ be a parameter and $f : G \to \mathbb{C}$ is a function. The η-large spectrum of f is the level set of $\widehat{G}$ defined by*

$$\mathrm{Spec}_\eta(f) = \{\, |\widehat{f}| \geqslant \eta \|f\|_{L^1} \}.$$

We also need to recall one piece of terminology from [82, Section 4], which is only used in this section. Write $\mathbb{D}$ for the unit disk, and let μ be any measure on G. Given a parameter $\theta \in (0,1]$, we say that a set Λ of characters is (θ,μ)-dissociated when, for every function $\omega : \Lambda \to \mathbb{D}$, we have

$$\int \prod_{\lambda \in \Lambda} \Big(1 + \mathrm{Re}[\omega(\lambda)\lambda]\Big) \mathrm{d}\mu \leqslant e^\theta,$$

and when $\theta = 1$ we simply say that Λ is μ-dissociated. We may now quote two lemmas of local spectral analysis from [82], with minor tweaks in both cases.

LEMMA 5.3 (Local Chang bound). *Let $\eta \in (0,1]$ be a parameter, and suppose that B is a subset of G and X is a subset of B of density τ. Then every μ_B-dissociated subset of $\mathrm{Spec}_\eta(\mu_X)$ has size at most $C\eta^{-2}\log\tau^{-1}$.*

PROOF. This is [82, Lemma 4.6], specialized to the case where $f = \mu_X$ and $\mu = \mu_B$, so that with the notation from there $L_{\mu_X,\mu_B} = \tau^{-1/2}$. □

LEMMA 5.4 (Annihilating locally dissociated sets). *Let $\nu \in (0,1]$ be a parameter. Suppose that $\mathcal{B}$ is a regular Bourgain system, Δ is a set of characters, and m is the size of the largest μ_B-dissociated subset of Δ, or 1 if there is no such subset. Then*

there exists a Bohr set $\widetilde{B}$ *of dimension at most* m *and radius equal to* c/m *such that* Δ *is* ν*-annihilated by* $B_{c\nu/d^2m} \cap \widetilde{B}_\nu$.

PROOF. This is [82, Lemma 6.3] with $\eta = 1$ and $m = \max(k, 1)$, and two minor tweaks: $\mathcal{B}$ is a Bourgain system instead of a Bohr set and a few changes of variables have been effected. Since the proof requires only a regularity estimate of the type of Lemma 4.12, the generalization to Bourgain systems is immediate. □

As usual these two ingredients combine to show that the large spectrum of a dense subset of a Bourgain system may be efficiently annihilated. Before carrying this out, we introduce a last definition which serves to simplify our technical statements.

DEFINITION 5.5. *Let* $m \geqslant 1$ *be a parameter and suppose that* $\mathcal{B}$ *is a Bourgain system. We say that* $\mathcal{B}$ *is* m*-controlled when it has dimension at most* m *and density at least* $\exp[-Cm \log m]$.

We are now ready to introduce the main technical tool of this paper. Recall that $\ell(x)$ stands for $\log(e/x)$ here and throughout the article.

PROPOSITION 5.6 (Local spectrum annihilation). *Let* $\eta, \nu \in (0, 1]$ *be parameters. Suppose that* $\mathcal{B}$ *is a regular Bourgain system and* X *is a subset of* B *of relative density* τ. *Then* $\mathrm{Spec}_\eta(\mu_X)$ *is* ν*-annihilated by a regular Bourgain system of the form*

$$\mathcal{B}_{c\nu/d^2m} \wedge \widetilde{\mathcal{B}}_\nu \quad \textit{where} \quad m \leqslant C\eta^{-2}\ell(\tau)$$

and $\widetilde{\mathcal{B}}$ *is an* m*-controlled Bourgain system.*

PROOF. Let m denote the size of the largest μ_B-dissociated subset of $\mathrm{Spec}_\eta(\mu_X)$, or 1 when there is no such set. By Lemma 5.3, we have $m \leqslant C\eta^{-2}\ell(\tau)$. By Lemma 5.4, we also know that $\mathrm{Spec}_\eta(\mu_X)$ is ν-annihilated by a regular Bourgain system $\overline{\mathcal{B}} := \mathcal{B}_{c\nu/d^2m} \wedge \widetilde{\mathcal{B}}_\nu$, where $\widetilde{\mathcal{B}}$ is the Bourgain system induced by a Bohr set

of dimension $d \leqslant m$ and radius $\delta = c/m$. By Lemma 4.11, we may further ensure that $\overline{\mathcal{B}}$ is regular up to dilating it by a factor $\asymp 1$, which does not affect the shape of the above intersection except in the value of the constants. By Lemma 4.3, we also see that $\widetilde{\mathcal{B}}$ has dimension at most $6m$ and density at least $\exp[-Cm\log m]$, so that the result follows by replacing $6m$ with m and adapting the constants. □

6. Roth's theorem for Bourgain systems

This section is concerned with a local version of Roth's theorem [69], first considered by Sanders [80], which applies to dense subsets of a Bourgain system. Since the pioneering work of Bourgain [5], modern proofs of Roth's theorem [81,82] all share the same global structure and proceed by an iteration on subsets of Bohr sets. An important observation made in [80] is that this iteration may be initialized inside a certain Bohr set instead of the whole group, and further that one may perform the same iteration on Bourgain systems in place of Bohr sets.

However the quantitative estimates obtained in [80] correspond roughly in strength to a range of $\alpha \gtrsim (\log N)^{-1/3}$ in Roth's theorem, while the best-known range, also by Sanders [81], is now $\alpha \gtrsim (\log N)^{-1}$. Conceptually, there is no obstacle in obtaining this better quantitative dependency with Bourgain systems, and for the same local initialization, however on a technical level it is not entirely straightforward as most density-increment statements then take a different shape. We carry out this process in this section; since it is not the right place here to present the whole argument of [81], we only include the main structural results we need from it and indicate the changes that need to be done to other. Unfortunately, this means that the reader needs either to be conversant with [81], or to read this section conditionally on Proposition 6.4 below. What we obtain eventually is the following quantitative improvement of [80, Theorem 5.1].

PROPOSITION 6.1 (Local Sanders-Roth theorem). *Suppose that $\mathcal{B}$ is a regular Bourgain system and A is a subset of B of relative density α such that $A - A$*

contains no element of order 2. *Then*

$$\langle 1_A * 1_A, 1_{2\cdot A}\rangle_{L^2} \geqslant \exp\Big[-C(\alpha^{-1}+d)\ell(\alpha)^6\ell(\alpha/d)\Big]\cdot b^2.$$

We make a brief comment here on the shape of the above proposition. The three-term arithmetic progressions contained in a set A are precisely the triples (x, y, z) of A^3 such that $x + z = 2 \cdot y$. The assumption on A shows that the change of variables $y \mapsto 2 \cdot y$ is injective on A, from which we see that the total number of such progressions is equal to $\langle 1_A * 1_A, 1_{2\cdot A}\rangle_{L^2} \cdot |G|^2$. We invite the reader to keep this observation in mind, as it is used implicitely in later arguments.

We now present our modified version of the argument of [81]. To begin with, we reconstitute the L^2 density-increment strategy entirely as it takes a different form for Bourgain systems, which determines the shape of iterative statements. The following lemma is the usual argument that allows one to pass from large energy of the Fourier transform over a character set, to a density increment on any set annihilating those characters.

LEMMA 6.2. *Let $\rho, \kappa \in (0, 1]$ be parameters. Suppose that $\mathcal{B}$ is a regular Bourgain system, A is a subset of B of relative density α, T is a subset of B_ρ and Δ is a set of characters. Assume also that $\rho \leqslant c\kappa\alpha/d$ and write $f_A = 1_A - \alpha 1_B$. Then if*

$$\sum_{\Delta} |\widehat{f_A}|^2 \geqslant \kappa\alpha^2 b \qquad \text{and } \Delta \text{ is } \tfrac{1}{2}\text{-annihilated by } T,$$

we have $\|1_A * \mu_T\|_\infty \geqslant (1 + 2^{-3}\kappa)\alpha$.

PROOF. For every character $\gamma \in \Delta$ we know that $|1 - \gamma| \leqslant 1/2$ on T, and therefore $|\widehat{\mu_T}(\gamma) - 1| \leqslant \mathbb{E}_T|1 - \gamma| \leqslant \frac{1}{2}$ and $|\widehat{\mu_T}(\gamma)| \geqslant \frac{1}{2}$. Inserting this into the energy lower bound, we have, via Parseval,

$$\begin{aligned}\tfrac{1}{4}\kappa\alpha^2 b &\leqslant \textstyle\sum_{\widehat{G}} |\widehat{f_A}|^2|\widehat{\mu}_T|^2\\ &= \langle f_A * \mu_T, f_A * \mu_T\rangle_{L^2}.\end{aligned}$$

Expanding this scalar product, and with the help of Lemma 4.12, we obtain

$$\begin{aligned}\tfrac{1}{4}\kappa\alpha^2 b &\leqslant \|1_A * \mu_T\|_2^2 - 2\alpha \langle 1_A * \mu_T, 1_B * \mu_T\rangle_{L^2} + \alpha^2 \langle 1_B * \mu_T, 1_B * \mu_T\rangle_{L^2} \\ &= \|1_A * \mu_T\|_2^2 - 2\alpha b \langle 1_A, \mu_B * \mu_T * \mu_{-T}\rangle_{L^2} + \alpha^2 b \langle 1_B, \mu_B * \mu_T * \mu_{-T}\rangle_{L^2} \\ &= \|1_A * \mu_T\|_2^2 - \Big(1 + O\Big(\tfrac{\rho d}{\alpha}\Big)\Big)\alpha^2 b.\end{aligned}$$

Choosing $\rho \leqslant c\kappa\alpha/d$, we have then

$$\begin{aligned}(1 + 2^{-3}\kappa)\alpha^2 b &\leqslant \|1_A * \mu_T\|_2^2 \\ &\leqslant \|1_A * \mu_T\|_\infty \|1_A * \mu_T\|_1 \\ &= \|1_A * \mu_T\|_\infty \cdot \alpha b.\end{aligned}$$

Dividing both sides by αb concludes the proof. □

As usual this may be combined with a statement on the local annihilation of the large spectrum, such as Proposition 5.6, to recover an L^2-density increment lemma.

PROPOSITION 6.3 (L^2 density-increment). *Let $\kappa, \eta \in (0, 1]$ be parameters. Suppose that $\mathcal{B}$, $\dot{\mathcal{B}}$ are Bourgain systems and $\mathcal{B}$ is regular, A is a subset of B of relative density α and X is a subset of $\dot{B}$ of relative density τ. Assume also that $\dot{\mathcal{B}} \leqslant_\rho \mathcal{B}$ with $\rho \leqslant c\kappa\alpha/d$ and write $f_A = 1_A - \alpha 1_B$. Then if*

$$\sum_{\mathrm{Spec}_\eta(\mu_X)} |\widehat{f_A}|^2 \geqslant \kappa\alpha^2 b,$$

there exists an m-controlled Bourgain system $\widetilde{\mathcal{B}}$ such that

$$\begin{aligned}\overline{\mathcal{B}} &= \dot{\mathcal{B}}_{c/\dot{d}^2 m} \wedge \widetilde{\mathcal{B}} \quad \text{is regular,} \\ m &\leqslant C\eta^{-2}\ell(\tau), \\ \|1_A * \mu_{\overline{B}}\|_\infty &\geqslant (1 + 2^{-3}\kappa)\alpha.\end{aligned}$$

PROOF. By Proposition 5.6, $\mathrm{Spec}_\eta(\mu_X)$ is $\frac{1}{2}$-annihilated by a regular Bourgain system of the form $\overline{\mathcal{B}} = \dot{\mathcal{B}}_{cd^2/m} \wedge \widetilde{\mathcal{B}}$, where $\widetilde{\mathcal{B}} = \widetilde{\mathcal{B}}'_{1/2}$ and $\widetilde{\mathcal{B}}'$ is an m'-controlled Bourgain system with $m' \leqslant C\eta^{-2}\ell(\tau)$. Note that by Lemma 4.6, $\widetilde{\mathcal{B}}$ is $O(m')$-controlled. Applying then Lemma 6.2 with $\Delta = \mathrm{Spec}_\eta(\mu_X)$ and $T = \overline{\mathcal{B}} \leqslant \dot{\mathcal{B}}$ concludes the proof. □

We now take a big step forward and claim that the following analog of [81, Lemma 6.2] holds. This involves a careful examination of the argument of [81], and we regret imposing the double-checking process below on the reader, however past this point our argument is again self-contained.

PROPOSITION 6.4 (Iterative lemma on two scales). *Suppose that $\mathcal{B}$, $\mathcal{B}'$ are regular Bourgain systems, A is a subset of B of relative density α and A' is a subset of B' of relative density α'. Assume also that $\mathcal{B}' \leqslant_\rho \mathcal{B}$ with $\rho \leqslant c\alpha/d$. Then either*

(i) *(Many three-term arithmetic progressions)*

$$\langle 1_A * 1_{A'}, 1_{-A}\rangle_{L^2} \geqslant \exp\Big[-C\alpha^{-1}\ell(\alpha') - Cd'\ell(\alpha'/d')\Big]bb',$$

(ii) *(Density increment)*

there exists an m-controlled Bourgain system $\widetilde{\mathcal{B}}$ with

$$\overline{\mathcal{B}} = \mathcal{B}'_{(\alpha\alpha'/2d')^C} \wedge \widetilde{\mathcal{B}} \quad \textit{regular},$$
$$m \leqslant C\alpha^{-1}\ell(\alpha)^3\ell(\alpha'),$$
$$\|1_A * \mu_{\overline{B}}\|_\infty \geqslant (1 + 2^{-13})\alpha.$$

PROOF. This is obtained by replacing each occurence of the energy-increment lemma [81, Lemma 3.8] for Bohr sets by its Bourgain system counterpart, viz. Proposition 6.3. Essentially two types of L^2 density-increment appear in Sanders' argument, and we now describe them, using the notation of Proposition 6.3. In every application of [81, Lemma 3.8] the Bourgain system $\dot{\mathcal{B}}$ is (eventually) a dilate of the

Bourgain system $\mathcal{B}$ by a factor $(\alpha\alpha'/2d')^{O(1)}$, and therefore we only need determine the parameters κ, η, τ.

The first type of L^2 density-increment appears in the proof of [**81**, Lemma 4.2] on p. 626 with parameters $\kappa \asymp 1$, $\eta \asymp \alpha^{1/2}$, $\tau \gg \alpha'$, so that $m \leqslant C\alpha^{-1}\ell(\alpha')$ upon applying Proposition 6.3. The same density-increment is featured in [**81**, Proposition 4.1] which is just an iteration of the previous lemma.

A second type of density-increment arises in the proof of [**81**, Corollary 5.2] on pp. 630–632 which involves certain densities σ and λ, and which features parameters $\kappa \asymp \lambda$, $\eta \asymp 1$,

$$\tau \geqslant \exp[-C\lambda^{-2}\ell(\sigma)\ell(\lambda\alpha)^2\ell(\alpha)] \quad \text{so that} \quad m \leqslant C\lambda^{-2}\ell(\sigma)\ell(\lambda\alpha)^2\ell(\alpha)$$

upon applying Proposition 6.3. This is finally combined with [**81**, Proposition 4.1] on p. 633 to obtain [**81**, Lemma 6.2], to the effect that we either have an L^2 density-increment of the first type, or of the second type with $\lambda \asymp 1$ and $\sigma \geqslant \exp[-C\alpha^{-1}\ell(\alpha')]$, and therefore such that $\kappa \asymp 1$ and $m \leqslant C\alpha^{-1}\ell(\alpha)^3\ell(\alpha')$ in the application of Proposition 6.3. Choosing $\mathcal{B}'' = \mathcal{B}'_{c\alpha'/d'}$ in (the Bourgain system version of) [**81**, Lemma 6.2] and using Lemma 4.6, we obtain an alternative case (i) of the desired shape.

Since, by Lemma 4.12, Bourgain systems satisfy the same regularity estimates as Bohr sets, we may replace the latter by the former and apply Proposition 6.3 everywhere as claimed, thereby obtaining the desired iterative lemma. Finally, the constant 2^{-13} may be extracted from [**81**] although its precise value is unimportant; it is just convenient to write down an explicit value for later computations. □

At this point we recall a simple technique, originating in Bourgain's proof of Roth's theorem [**5**, (5.13)–(5.18)], which allows one to pass from two scales to one in iterative statements.

LEMMA 6.5. *Let $\theta \in (0, 1]$ be a parameter. Suppose that $\mathcal{B}$, $\mathcal{B}'$, $\mathcal{B}''$ are Bourgain systems, $\mathcal{B}$ is regular and A is a subset of B of relative density α. Assume also that*

$\mathcal{B}' \leqslant_\rho \mathcal{B}$ and $\mathcal{B}'' \leqslant_\rho \mathcal{B}$ with $\rho \leqslant c\theta\alpha/d$. Then either

$$\max\left(\|1_A * \mu_{B'}\|_\infty, \|1_A * \mu_{B''}\|_\infty\right) \geqslant \left(1 + \tfrac{\theta}{2}\right)\alpha$$

or there exists x such that $1_A * \mu_{B'}(x) \geqslant (1-\theta)\alpha$ and $1_A * \mu_{B''}(x) \geqslant (1-\theta)\alpha$.

PROOF. A quick regularity computation via Lemma 4.12 yields

$$\begin{aligned}\mathbb{E}_B(1_A * \mu_{B'} + 1_A * \mu_{B''}) &= \langle 1_A, \mu_B * \mu_{B'}\rangle + \langle 1_A, \mu_B * \mu_{B''}\rangle \\ &= 2\alpha + O(\rho d) \\ &\geqslant (2 - \tfrac{\theta}{2})\alpha\end{aligned}$$

provided that $\rho \leqslant c\theta\alpha/d$. By the pigeonhole principle, there exists $x \in G$ such that

$$1_A * \mu_{B'}(x) + 1_A * \mu_{B''}(x) \geqslant (2 - \tfrac{\theta}{2})\alpha.$$

Assuming that we are not in the first case of the lemma, we have

$$1_A * \mu_{B'}(x) \geqslant (2 - \tfrac{\theta}{2})\alpha - (1 + \tfrac{\theta}{2})\alpha = (1-\theta)\alpha$$

and similarly for $1_A * \mu_{B''}(x)$. □

With this technique in hand, we may modify Proposition 6.4 so as to make the iteration easier to perform. Once this is done, Proposition 6.1 is derived by a standard, yet computationally intensive iterative process. For this argument to work however, we need to make the assumption that the set A contains no degenerate arithmetic progressions at each step of the iteration.

PROPOSITION 6.6 (Final iterative lemma). *Suppose that G has odd order, $\mathcal{B}$ is a regular Bourgain system, and A is a subset of B of relative density α such that $A - A$ contains no element of order 2. Then either*

(i) *(Many three-term arithmetic progressions)*

$$\langle 1_A * 1_A, 1_{2\cdot A}\rangle_{L^2} \geqslant \exp\Big[-C\alpha^{-1}\ell(\alpha) - Cd\ell(\alpha/d)\Big]\cdot b^2,$$

(ii) *(Density increment)*

there exist Bourgain systems $\widehat{\mathcal{B}}$, $\widetilde{\mathcal{B}}$ *and an element* $u \in \{1,-2\}$ *such that*

$$\begin{aligned}
&\overline{\mathcal{B}} = \widehat{\mathcal{B}} \wedge \widetilde{\mathcal{B}} \quad \textit{is regular},\\
&\widehat{\mathcal{B}} = u\cdot \mathcal{B}_{(\alpha/2d)^C}, && \widehat{b} \geqslant \exp\Big[-Cd\ell(\alpha/d)\Big]\cdot b,\\
&\widetilde{d} \leqslant C\alpha^{-1}\ell(\alpha)^4, && \widetilde{b} \geqslant \exp[-C\alpha^{-1}\ell(\alpha)^5],\\
&\|1_A * \mu_{\overline{B}}\|_\infty \geqslant (1+2^{-16})\alpha.
\end{aligned}$$

PROOF. Let $\theta = 2^{-15}$ and define regular Bourgain systems $\mathcal{B}' = \mathcal{B}_{c\alpha/d}$ and $\mathcal{B}'' = \mathcal{B}'_{c'\alpha/d}$ with the help of Lemma 4.11. Now apply Lemma 6.5 to A and $\mathcal{B}, \mathcal{B}', \mathcal{B}''$: in the first case of that lemma, we are in the second case of the proposition, while in the second case we may find an element x such that $A' \coloneqq (A-x)\cap B'$ has relative density $\alpha' \geqslant (1-2^{-15})\alpha$ in B', and $A'' \coloneqq (A-x)\cap B''$ has relative density at least $\frac{1}{2}\alpha$ in B''; the latter weak bound suffices for our purposes.

We let $\widehat{A}'' = -2\cdot A''$ and $\widehat{\mathcal{B}}'' = -2\cdot \mathcal{B}''$, so that from the injectivity of $y \mapsto 2\cdot y$ on A'' and the bound $|\widehat{B}''| \leqslant |B''|$, we deduce that $\widehat{A}''$ has density at least $\frac{1}{2}\alpha$ in $\widehat{B}''$. Furthermore, by Lemma 4.9, we see that $\widehat{\mathcal{B}}''$ is a Bourgain system of dimension at most d'' and, since $\widehat{B}''$ contains $\widehat{A}''$, of density at least $\frac{1}{2}\alpha b''$. Observe finally that with these choices of A' and $\widehat{A}''$, we have

$$\langle 1_A * 1_A, 1_{2\cdot A}\rangle_{L^2} = \langle 1_{A-x} * 1_{2x-2\cdot A}, 1_{x-A}\rangle_{L^2} \geqslant \langle 1_{A'} * 1_{\widehat{A}''}, 1_{-A'}\rangle_{L^2}. \tag{6.1}$$

We now apply Proposition 6.4 to the sets A' and $\widehat{A}''$, located respectively in B' and $\widehat{B}''$. In the first case of that proposition, it follows from (6.1) and Lemma 4.6 that we are in the first case of the proposition we seek to prove. In the second case

of Proposition 6.4, we obtain a regular Bourgain system $\overline{\mathcal{B}} = \widehat{\mathcal{B}} \wedge \widetilde{\mathcal{B}}$ where

$$\widehat{\mathcal{B}} = (-2 \cdot \mathcal{B}'')_{(\alpha/2d)^C} = -2 \cdot \mathcal{B}''_{(\alpha/2d)^C} = -2 \cdot \mathcal{B}_{(\alpha/2d)^{C'}}$$

and $\widetilde{\mathcal{B}}$ is $C\alpha^{-1}\ell(\alpha)^4$-controlled, and such that

$$\|1_A * \mu_{\overline{B}}\|_\infty \geqslant \|1_{A'} * \mu_{\overline{B}}\|_\infty \geqslant (1 + 2^{-13})\alpha' \geqslant (1 + 2^{-14})\alpha.$$

Applying Lemma 4.6 to $\widehat{\mathcal{B}} = \widehat{\mathcal{B}}''_{(\alpha/2d)^C}$, recalling that $\widehat{b}'' \geqslant \frac{1}{2}\alpha b''$, and via Definition 5.5, we conclude that we are in the second case of the proposition that we intend to prove. □

Proof of Proposition 6.1. We construct iteratively sequences of subsets A_i of regular Bourgain systems $\mathcal{B}^{(i)}$ of density α_i, such that A_i is contained in a translate of A. Since $A_i - A_i$ is a subset $A - A$, it does not contain any element of order 2 either. We initiate the iteration with $A_1 = A$ and $\mathcal{B}^{(1)} = \mathcal{B}$.

At each step we apply Proposition 6.6 to the set A_i, and in the first case of that proposition we stop the iteration, while in the second case we let $\mathcal{B}^{(i+1)} = \overline{\mathcal{B}}^{(i)}$ with the notation from there, and we pick x_i and $A_{i+1} = (A_i - x_i) \cap \overline{B}^{(i)}$ so that A_{i+1} has relative density $\alpha_{i+1} = \|1_{A_i} * \mu_{\overline{B}^{(i)}}\|_\infty$ in $\overline{B}^{(i)}$.

Since $\alpha_{i+1} \geqslant (1 + c)\alpha_i$ whenever A_{i+1} is defined, the iteration proceeds for a number of steps bounded by $C\ell(\alpha)$. At each step, we obtain Bourgain systems $\widehat{\mathcal{B}}^{(i)}$ and $\widetilde{\mathcal{B}}^{(i)}$ and an element $u_i \in \{1, -2\}$ such that

$$\mathcal{B}^{(i+1)} = \widehat{\mathcal{B}}^{(i)} \wedge \widetilde{\mathcal{B}}^{(i)} \quad \text{is regular,} \tag{6.2}$$

and, since $\alpha_i \geqslant \alpha$, such that

$$\widehat{\mathcal{B}}^{(i)} = u_i \cdot \mathcal{B}^{(i)}_{(\alpha_i/2d_i)^C}, \qquad \widehat{b}_i \geqslant \exp\Big[-Cd_i\ell(\alpha/d_i)\Big] \cdot b_i, \tag{6.3}$$

$$\widetilde{d}_i \leqslant C\alpha^{-1}\ell(\alpha)^4, \qquad \widetilde{b}_i \geqslant \exp\Big[-C\alpha^{-1}\ell(\alpha)^5\Big]. \tag{6.4}$$

Iterating $i-1$ times (6.2) and (6.3), we obtain a Bourgain system of the form

$$\mathcal{B}^{(i)} = \widetilde{\mathcal{B}}^{(i-1)} \wedge u_{i-1} \cdot \Big(\dots u_2 \cdot (\widetilde{\mathcal{B}}_*^{(1)} \wedge u_1 \cdot \widetilde{\mathcal{B}}_*) \dots \Big)$$

where the stars stand for certain dilations. This is not exactly an intersection of Bourgain systems, however the argument used in the proof of Lemma 4.8 is easily adapted to show that $\mathcal{B}^{(i)}$ has dimension at most

$$d_i \leqslant 2(d + \widetilde{d}_1 + \cdots + \widetilde{d}_{i-1}).$$

By (6.4) and since $i \leqslant C\ell(\alpha)$, this yields $d_i \leqslant 2d + C\alpha^{-1}\ell(\alpha)^5$.

Applying Lemma 4.8 to the intersection (6.2), and with (6.3) and (6.4), we also obtain

$$\begin{aligned} b_{i+1} &\geqslant 4^{-(\widehat{d}_i + \widetilde{d}_i)} \cdot \widehat{b}_i \cdot \widetilde{b}_i \\ &\geqslant \exp\Big[- C(\alpha^{-1} + d)\ell(\alpha)^5 \ell(\alpha/d) \Big] \cdot b_i. \end{aligned}$$

Iterating this at most $C\ell(\alpha)$ times, we obtain

$$b_i \geqslant \exp\Big[- C(\alpha^{-1} + d)\ell(\alpha)^6 \ell(\alpha/d) \Big] \cdot b.$$

When the algorithm stops, we have therefore

$$\langle 1_{A_i} * 1_{A_i}, 1_{2\cdot A_i} \rangle_{L^2} \geqslant \exp\Big[- C\alpha^{-1}\ell(\alpha) - Cd_i\ell(\alpha/d_i) \Big] \cdot b_i^2.$$

Inserting the bounds on d_i and b_i in the above, and recalling that A_i is contained in a translate of A, this concludes the proof.

7. From small doubling to three-term arithmetic progressions

This section is concerned with the proof of Theorem 1.3 and the related Corollary 1.6. As mentioned before, an extremely important tool for us is the recent correlation-based Bogolyubov-Ruzsa lemma of Sanders [**83**]. In our situation, it serves to pass from a set of small doubling to one with high density in a coset

progression, which is a particular type of Bourgain system. The local Sanders-Roth theorem of the previous section may then be applied to this new set, to show that it contains a nontrivial three-term arithmetic progression; this is the main observation of this paper. We now quote the main result of [83], with a minor tweak to ensure regularity.

PROPOSITION 7.1 (Correlation Bogolyubov-Ruzsa lemma [83]). *Let $K \geqslant 1$ be a parameter, and suppose that A is a subset of G such that $|A+A| \leqslant K|A|$. Then there exists a d-dimensional coset progression M inducing a regular Bourgain system and such that*

$$\begin{aligned} \|1_A * \mu_M\|_\infty &\geqslant \tfrac{1}{2K}, \\ d &\leqslant C(\log K)^6, \\ |M| &\geqslant \exp\Big[-C(\log K)^6(\log\log K)\Big] \cdot |A|. \end{aligned}$$

PROOF. Without the regularity condition, this is [83, Theorem 10.1] with $A = S$ and $\varepsilon = \frac{1}{2}$. To obtain regularity, one may simply follow the proof in [83], stopping just before the application of [83, Lemma 10.2], and dilating by a certain constant factor the coset progression M obtained at this point. By Lemmas 4.6 and 4.11, one may choose this constant so that the dilated induced Bourgain system is regular, while losing at most a factor $e^{-C(\log K)^6}$ in size, and the rest of the proof goes unchanged. □

It is crucial for our argument that this statement makes no assumption of density on the set A, whereas the earlier Bogolyubov-Chang-type lemma [80, Proposition 6.1] used by Sanders does. In terms of bounds, we could also allow for $d \leqslant K^{1+o(1)}$ and $|M| \geqslant e^{-CK^{1+o(1)}}|A|$ in Proposition 7.1, without affecting the quality of bounds in Theorem 1.3; however we do not know of any argument significantly simpler than that of [83] to obtain such estimates.

We now present the proof of Theorem 1.3, following the usual approach of estimating the total number of three-term arithmetic progressions, only to compare it later to the number of trivial ones. Corollary 1.6 then follows by inserting the bound of Theorem 1.3 into the argument of [80].

PROPOSITION 7.2. *Let $K \geqslant 1$ be a parameter. Suppose that A is a subset of G such that $|A+A| \leqslant K|A|$ and $A-A$ contains no element of order 2. Then*

$$\langle 1_A * 1_A, 1_{2\cdot A}\rangle_{L^2} \geqslant \exp\left[-CK(\log K)^7\right] \cdot \mu_G(A)^2.$$

PROOF. Let M be the coset progression given by Proposition 7.1, and write $\mathcal{M}$ for its induced regular Bourgain system. By the correlation conclusion, we may pick an element x such that $A' = (A-x) \cap M$ has relative density $\frac{1}{2K}$ in M. Applying then Proposition 6.1 to A' and $\mathcal{M}$, we obtain

$$\langle 1_A * 1_A, 1_{2\cdot A}\rangle_{L^2} \geqslant \langle 1_{A'} * 1_{A'}, 1_{2\cdot A'}\rangle_{L^2} \geqslant \exp[-C(K+d)(\log K)^6(\log Kd)] \cdot \mu_G(M)^2.$$

This yields the desired estimate upon inserting the bounds from Proposition 7.1. □

Proof of Theorem 1.3. Write $K = |A+A|/|A|$. If $A-A$ contains an element $x-y$ of order 2, we readily find a nontrivial, degenerate arithmetic progression (x, y, x) in A. Otherwise, Proposition 7.2 tells us that A possesses at least $e^{-CK(\log K)^7}|A|^2$ three-term arithmetic progressions, while the number of trivial ones is at most $|A|$. By the assumption on K, we are then ensured to find at least one nontrivial arithmetic progression in A. □

Proof of Corollary 1.6. It suffices to insert the bounds of Theorem 1.3 in the proof of [80, Theorem 1.5] on pp. 230–231. □

8. From small doubling to long arithmetic progressions

In this section we derive Theorem 1.5, basing ourselves on the approach of Croot et al. [9], which divides roughly into three steps. In the first step, one produces a large, structured set of almost periods of the convolution of the set A under

consideration with itself. The second step is to show, by a packing argument, that the set $A+A$ necessarily contains a translated copy of subset of this set of almost-periods of a certain size. The third step is to pick such a subset with basic additive structure, such as an arithmetic progression.

The original argument of [9] is based on Ruzsa's modelling lemma [75], which has no efficient equivalent for general abelian groups, and therefore we need to use again a modelling approach based on the Bogolyubov-Ruzsa lemma of Sanders. In contrast with the previous section however, we now need a version of this lemma that provides us with a containment conclusion, and for this we quote [83, Theorem 1.1].

PROPOSITION 8.1 (Containment Bogolyubov-Ruzsa lemma [83]). *Let $K \geqslant 1$ be a parameter, and suppose that A is a subset of G such that $|A+A| \leqslant K|A|$. Then there exists a d-dimensional coset progression M contained in $2A-2A$ and such that*

$$d \leqslant C(\log K)^6 \quad \text{and} \quad |M| \geqslant \exp\Big[-C(\log K)^6(\log\log K)\Big] \cdot |A|.$$

As noted in [83, Section 3], this version can be deduced from Proposition 7.1. The containment conclusion is sufficient in our situation, because the Croot-Sisask lemma works under a doubling hypothesis, whereas the iterative argument used in the proof of Roth's theorem requires an assumption of density instead. Our reason for emphasizing this point is that the containment version above is easier to obtain than the correlation one, and is explained in depth in a survey by Sanders [84]. Although the type of structure obtained there is different, consisting of a convex coset progression instead, this would not affect our argument much since this object is also a Bourgain system, as can be seen from [84, Section 4].

We now proceed to the proof, starting with the following lemma which serves to collect together certain computations from [9] on L^p and $L^{p/2}$ norms of convolutions.

LEMMA 8.2. *Let $p \geqslant 2$ and $K \geqslant 1$ be parameters. Suppose that A is a subset of G such that $|A+A| \leqslant K|A|$. Then*

$$\mu_G(A+A)^{1/p} \leqslant K^{1/2}\|1_A * \mu_A\|_{p/2}^{1/2} \quad \textit{and} \quad \|1_A * \mu_A\|_{p/2}^{1/2} \leqslant K^{1/2}\|1_A * \mu_A\|_p.$$

PROOF. By Hölder's inequality we have

$$\mu_G(A) = \mathbb{E}_G 1_A * \mu_A \leqslant \mu_G(A+A)^{1-2/p}\|1_A * \mu_A\|_{p/2},$$

from which the first estimate follows upon rearranging and taking square roots. To obtain the second, apply Cauchy-Schwarz and the first estimate in

$$\left[\mathbb{E}_G\,(1_A * \mu_A)^{p/2}\right]^2 \leqslant \mu_G(A+A)\|1_A * \mu_A\|_p^p \leqslant K^{p/2}\|1_A * \mu_A\|_{p/2}^{p/2}\|1_A * \mu_A\|_p^p.$$

The result follows upon taking p-th roots, then dividing both sides by $\|1_A * \mu_A\|_{p/2}^{1/2}$. □

An important tool from [9] is a version of the Croot-Sisask lemma [11] that serves to smooth the convolution of two sets by an iterated convolution factor. The precise statement we need is a standard consequence of [9, Theorem 6.1]; an exposition of it by the author may be found in [53, Section 7].

LEMMA 8.3 (Croot-Sisask L^p-smoothing). *Let $K, L \geqslant 1$, $\theta \in (0, K^{-1/2}]$, $p \in 2\mathbb{N}$, $\ell \in \mathbb{N}$ be parameters. Suppose that A, S, T are subsets of G such that $|A+S| \leqslant K|A|$ and $|S+T| \leqslant L|S|$. Then there exists a subset X of T of size $|X| \geqslant (2L)^{-Cp\ell^2/\theta^2}|T|$ such that*

$$\|1_A * \mu_S - 1_A * \mu_S * \lambda_X^{(\ell)}\|_p \leqslant \theta\|1_A * \mu_S\|_{p/2}^{1/2}$$

where $\lambda_X = \mu_X * \mu_{-X}$.

As anticipated, our first step is to produce a set of almost-periods of the convolution of a small doubling set with itself. Following [9], this is done by first smoothing this convolution by the iterated convolution of a certain set X, with the

difference that this set is now localized to a Bourgain system, which is taken to be a coset progression later on. Via the Fourier transform, any set annihilating the large spectrum of X induces a set of almost-periods of the smoothed convolution, and via the results of Section 5, we may choose this annihilator to be a large Bourgain system. Here we make a small parenthesis on notation: throughout this section, $a \sim b$ stands for $b/2 \leqslant a \leqslant 2b$.

PROPOSITION 8.4. *Let $K \geqslant 1$ and $p \in 2\mathbb{N}$ be parameters. Suppose that $\mathcal{B}$ is a regular Bourgain system and A is a subset of G such that $|A+A| \leqslant K|A|$ and $B \subset 2A-2A$. Then there exist $m \geqslant 1$ and Bourgain systems $\overline{\mathcal{B}}, \widetilde{\mathcal{B}}$ such that $\widetilde{\mathcal{B}}$ is m-controlled and*

$$\overline{\mathcal{B}} = \mathcal{B}_{c/(Kd^2m)} \wedge \widetilde{\mathcal{B}}_{c/K},$$
$$m \leqslant CpK(\log K)^3,$$

and for every $x \in \overline{B}$,

$$\|1_A * \mu_A - \tau_x 1_A * \mu_A\|_p \leqslant \tfrac{1}{2}\|1_A * \mu_A\|_p.$$

PROOF. First observe that, by the Plünnecke-Ruzsa-Petridis inequality [66],

$$|A+B| \leqslant |3A-2A| \leqslant K^5|A|,$$

and therefore we may apply Lemma 8.3 with $(S,T) = (A,B)$ and $L = K^5$, for parameters θ and ℓ to be determined later. This yields a subset X of B of relative density τ such that

$$\tau \geqslant \exp\Big[-Cp\ell^2\theta^{-2}\log K\Big], \tag{8.1}$$

$$\|1_A * \mu_A - 1_A * \mu_A * \lambda_X^{(\ell)}\|_p \leqslant \theta\|1_A * \mu_A\|_{p/2}^{1/2}. \tag{8.2}$$

We write I for the identity operator on functions, and given $x \in G$ we define the function $\widehat{x} : \widehat{G} \to G$ which maps γ to $\gamma(x)$. Consider now an arbitrary element

x of G, then by the triangle inequality and (8.2), we have

$$\begin{aligned}\|(I-\tau_x)1_A*\mu_A\|_p \leqslant\ & \|(I-\tau_x)(1_A*\mu_A-1_A*\mu_A*\lambda_X^{(\ell)})\|_p \\ & + \|1_{(A+A)\cup(A+A-x)}\cdot(I-\tau_x)1_A*\mu_A*\lambda_X^{(\ell)}\|_p \\ \leqslant\ & 2\theta\|1_A*\mu_A\|_{p/2}^{1/2}+2\mu_G(A+A)^{1/p}\|(I-\tau_x)1_A*\mu_A*\lambda_X^{(\ell)}\|_\infty.\end{aligned}$$

By Parseval, we have further

(8.3)
$$\|(I-\tau_x)1_A*\mu_A\|_p \leqslant 2\theta\|1_A*\mu_A\|_{p/2}^{1/2}+2\mu_G(A+A)^{1/p}\textstyle\sum_{\widehat{G}}|\widehat{1}_A||\widehat{\mu}_A||\widehat{\mu}_X|^{2\ell}|1-\widehat{x}|.$$

Invoking now Proposition 5.6 with a parameter $\nu\in(0,1]$, and recalling (8.1), we infer that $\mathrm{Spec}_{1/2}(\mu_X)$ is ν-annihilated by $\overline{\mathcal{B}}=\mathcal{B}_{c\nu/d^2m}\wedge\widetilde{\mathcal{B}}_\nu$, where $\widetilde{\mathcal{B}}$ is an m-controlled Bourgain system with $m\leqslant Cp\ell^2\theta^{-2}\log K$. From now on we restrict to $x\in\overline{B}$, so that, by considering separately the summation over $\mathrm{Spec}_{1/2}(\mu_X)$ in (8.3), we obtain

$$\|(I-\tau_x)1_A*\mu_A\|_p \leqslant 2\theta\,\|1_A*\mu_A\|_{p/2}^{1/2}+2(\nu+2^{1-2\ell})\,\mu_G(A+A)^{1/p}\textstyle\sum_{\widehat{G}}|\widehat{1}_A||\widehat{\mu}_A|.$$

By Parseval we know that $\sum_{\widehat{G}}|\widehat{1}_A||\widehat{\mu}_A|=1$. Applying finally Lemma 8.2, we obtain

$$\begin{aligned}\|(I-\tau_x)1_A*\mu_A\|_p &\leqslant \left(2\theta+2\nu K^{1/2}+2^{2-2\ell}K^{1/2}\right)\|1_A*\mu_A\|_{p/2}^{1/2} \\ &\leqslant \left(2\theta+2\nu K^{1/2}+2^{2-2\ell}K^{1/2}\right)K^{1/2}\|1_A*\mu_A\|_p.\end{aligned}$$

Choosing $\theta=K^{-1/2}/8$, $\nu=K^{-1}/16$ and $\ell\sim C\log K$, we obtain the desired L^p-estimate, and the bound on m follows by inserting the value of these parameters. □

Secondly, we need the following packing argument which may be extracted from the computations of [9], but whose proof we include for completeness. In practice we specialize f below to $1_A*\mu_A$ which has $A+A$ as support.

LEMMA 8.5. *Let $p \geqslant 2$ be a parameter. Suppose that $f : G \to \mathbb{C}$ and $R \subset G$ are such that, for all $t \in R$,*

$$\|(I - \tau_t)f\|_p \leqslant \tfrac{1}{2}\|f\|_p.$$

Then for every subset T of R of size $|T| < 2^p$, there exists a translate $x \in G$ such that $x + T \subset \mathrm{Supp}(f)$.

PROOF. Given a subset T of R, consider the quantity

$$I := \sum_{t \in T} \|f - \tau_t f\|_p^p,$$

so that by the assumptions of the lemma, we have at once $I \leqslant |T| \cdot 2^{-p}\|f\|_p^p$.

Now assume for contradiction that for every $x \in G$, the translate $x + T$ is not contained in $\mathrm{Supp}(f)$; then for every $x \in G$ we may find an element $t \in T$ such that $f(x + t) = 0$. Exchanging summations, this yields the lower bound

$$I = \mathbb{E}_G \textstyle\sum_{t \in T} |f - \tau_t f|^p \geqslant \mathbb{E}_G |f|^p.$$

Combining both bounds on I, we obtain

$$\|f\|_p^p \leqslant |T| 2^{-p} \|f\|_p^p.$$

We obtain a contradiction if $|T| < 2^p$, and therefore we find a translated copy of Y in the support of f in that case. □

Last, we need an analog for Bourgain systems in abelian groups of the well-known fact, used in [9], that Bohr sets of $\mathbb{Z}_N$ of radius δ and dimension d contain arithmetic progressions of length δN^d.

LEMMA 8.6. *Suppose that $\mathcal{B}$ is a Bourgain system of dimension d and $h \geqslant d$, and assume that $|B| \geqslant 2^{6h}$. Then there exists a subset T of B, which is either a proper arithmetic progression or a subgroup, of size $\frac{1}{4}|B|^{1/4h} \leqslant |T| \leqslant |B|^{1/2h}$.*

PROOF. Let $\eta = 2|B|^{-1/2h} \in (0, 2^{-2}]$ so that, by Lemma 4.6, we have

$$|B_\eta| \geqslant \exp\Big[\log|B| - d\log\tfrac{2}{\eta}\Big] \geqslant |B|^{1/2}.$$

Let $N = \lfloor \eta^{-1/2} \rfloor$, so that we have a sumset containment

$$N^2 B_\eta \subset B_{N^2\eta} \subset B. \tag{8.4}$$

Since $\eta^{-1/2} \geqslant 2$, we have also $\frac{1}{2}\eta^{-1/2} \leqslant N \leqslant \eta^{-1/2}$.

We are now in one of two cases. In the first, there exists an element x in B_η of order N, thus the arithmetic progression $T = [0, N-1]_{\mathbb{Z}} \cdot x$ is proper and, by (8.4), contained in B. Since $|T| = N$, we have also $\frac{1}{4}|B|^{1/4h} \leqslant |T| \leqslant |B|^{1/4h}$.

In the second case, every element of B_η has order at most N. Since $|B_\eta| \geqslant |B|^{1/2} \geqslant N$, we may pick $N-1$ distinct nonzero elements $x_1, \dots, x_{N-1} \in B_\eta$ and consider the subgroup T they generate, viz.

$$T = \langle x_1, \dots, x_{N-1}\rangle_{\mathbb{Z}} = [0, N-1]_{\mathbb{Z}} \cdot x_1 + \cdots + [0, N-1]_{\mathbb{Z}} \cdot x_{N-1}.$$

By (8.4) it follows again that T is contained in B, and the size of T satisfies

$$\tfrac{1}{4}|B|^{1/4h} \leqslant N \leqslant |T| \leqslant N^2 \leqslant |B|^{1/2h}.$$

□

We are now ready to combine the previous propositions into a proof of Theorem 1.5.

Proof of Theorem 1.5. By Proposition 8.1, we may find a d-dimensional coset progression $M \subset 2A - 2A$ such that

$$d \leqslant (\log K)^{O(1)} \quad \text{and} \quad |M| \geqslant \exp\Big[-(\log K)^{O(1)}\Big] \cdot |A|. \tag{8.5}$$

Up to dilating M by a constant factor, which preserves the above bounds by Lemma 4.6, we may assume via Lemma 4.11 that M induces a regular Bourgain system $\mathcal{M}$. By Lemma 4.5, that system also satisfies the dimension bound (8.5).

Applying now Proposition 8.4 with $\mathcal{B} = \mathcal{M}$ and a parameter $p \in 2\mathbb{N}$ to be determined later, we obtain Bourgain systems $\overline{\mathcal{B}}, \widetilde{\mathcal{B}}$ such that

$$\overline{\mathcal{B}} = \mathcal{M}_{(1/2dpK)^{O(1)}} \wedge \widetilde{\mathcal{B}}_{c/K}, \tag{8.6}$$

$$\widetilde{d} \leqslant CpK(\log K)^3, \tag{8.7}$$

$$\widetilde{b} \geqslant \exp\Big[-CpK(\log pK)(\log K)^3\Big], \tag{8.8}$$

where we have unfolded Definition 5.5, and such that

$$\|(I - \tau_x)1_A * \mu_A\| \leqslant \tfrac{1}{2}\|1_A * \mu_A\|_p \quad \text{for all} \quad x \in \overline{B}. \tag{8.9}$$

Applying Lemma 4.8 to the intersection (8.6), and considering (8.5) and (8.7), we obtain

$$\overline{d} \ll (\log K)^{O(1)} + pK(\log K)^3 \ll pK(\log K)^3$$

and we let $h = CpK(\log K)^3 \geqslant \overline{d}$. By Lemmas 4.6 and 4.8, we also obtain

$$\mu_G(\overline{B}) \geqslant \exp\Big[-Cd(\log dpK)\Big]\mu_G(M) \cdot \exp\Big[-C\widetilde{d}\log K\Big]\widetilde{b}$$

and therefore, by (8.5), (8.7) and (8.8),

$$|\overline{B}| \geqslant \exp\Big[-CpK(\log pK)(\log K)^3\Big] \cdot |A|.$$

Both the conditions $|\overline{B}| \geqslant |A|^{1/2}$ and $|\overline{B}| \geqslant 2^{6h}$ are satisfied provided

$$pK(\log pK)(\log K)^3 \leqslant c\log|A|. \tag{8.10}$$

Considering that $\overline{B} \subset M \subset 2A - 2A$, we thus have a rough estimate $|A|^{1/2} \leqslant |\overline{B}| \leqslant |A|^4$. By Lemma 8.6, we may therefore find a subset T of $\overline{B}$, which is either a

proper arithmetic progression or a subgroup, of size bounded by

$$\tfrac{1}{4}|A|^{1/8h} \leqslant \tfrac{1}{4}|\overline{B}|^{1/4h} \leqslant |T| \leqslant |\overline{B}|^{1/2h} \leqslant |A|^{2/h}.$$

Recalling our choice $h = CpK(\log K)^3$ and (8.10), this shows that

$$|T| = \exp\left[\Theta\Big(\frac{\log|A|}{pK(\log K)^3}\Big)\right].$$

The condition $|T| < 2^p$ is therefore satisfied if we choose

$$p \sim C\left(\frac{\log|A|}{K(\log K)^3}\right)^{1/2}.$$

It remains to check the conditions $p \geqslant 2$ and (8.10); these are seen to be satisfied for

$$K \leqslant \frac{c\log|A|}{(\log\log|A|)^5}$$

after a tedious, yet elementary computation. This yields the final size estimate

$$|T| = \exp\left[\Theta\Big(\frac{\log|A|}{K(\log K)^3}\Big)^{1/2}\right]$$

and since we verified the conditions $|T| < 2^p$ and (8.9), an application of Lemma 8.5 with $f = 1_A * \mu_A$ and $R = \overline{B}$ concludes the proof. □

9. Remarks

In this section we collect together certain remarks of expository or exploratory nature which have not found their way into the main text.

We first wish to explain in more detail how Theorem 1.1 follows from the results of the literature. Consider a set of integers A of doubling K, then for the purpose of finding arithmetic progressions in A, we may instead assume that A is a subset of a cyclic group of odd order of density $\gg K^{-4}$ and doubling K, via a partial Freiman isomorphy [**75**]. Applying [**80**, Proposition 6.1] to A, one obtains a regular Bohr set of dimension $d \ll K\log K$ and density $b \geqslant \exp[-CK(\log K)^2]$, on which

a certain translate of A has density $\gg K^{-1}$. In that setting, Proposition 6.1 of this article is just [**81**, Theorem 1.1], initializing the iterative argument from there on a Bohr set instead of the whole group; there is no need to consider Bourgain systems or 2-torsion. Proposition 6.1 thus specialized shows that A contains at least $\exp[-CK(\log K)^8] \cdot |A|^2$ three-term arithmetic progressions, and therefore at least one nontrivial progression for $K = |A+A|/|A|$ in the range specified by Theorem 1.1.

Secondly, we remark that the modelling argument used in Sections 7 and 8 could likely be adapted to other problems on dense sets, such as solving translation-invariant equations or finding long arithmetic progressions in $A + A + A$, to obtain a generalization of these results to the case of sets of small doubling in an arbitrary abelian group. However, it is not clear to the author whether it is worth pursuing such generalizations, given the current lack of combinatorial applications of the kind of Corollary 1.6 for results of this type.

Chapitre V. On systems of complexity one in the primes

Author: Kevin Henriot.

Abstract: Consider a translation-invariant system of linear equations $V\mathbf{x} = 0$ of complexity one, where V is an integer $r \times t$ matrix. We show that if A is a subset of the primes up to N of density at least $C(\log\log N)^{-1/25t}$, there exists a solution $\mathbf{x} \in A^t$ to $V\mathbf{x} = 0$ with distinct coordinates. This extends a quantitative result of Helfgott and de Roton for three-term arithmetic progressions, while the qualitative result is known to hold for all translation-invariant systems of finite complexity by the work of Green and Tao.

1. Introduction

Consider a matrix $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ with coefficients on each line summing to 0, a condition we term *translation-invariant*. We are interested in special instances of the problem of finding a distinct-coordinates solution $\mathbf{y} \in A^t$ to the system of equations $V\mathbf{y} = 0$, where A is a dense subset of the set $\mathcal{P}_N$ of the primes up to a large integer N, and when the relative density decays with N. Note that the distinct-coordinates condition excludes trivial solutions of the form $(u, \dots, u)$, while the conditions of homogeneity and translation-invariance on the system of equations are necessary to expect a Szemerédi-type theorem for $V\mathbf{y} = 0$, as can be seen by examining the case of a single linear equation (see e.g. [76, Theorem 1.3]).

We may assume that V has rank r up to removing redundant equations. Furthermore, we may work in practice with a parametrization $\psi : \mathbb{Z}^{t-r} \xrightarrow{\sim} \mathbb{Z}^t \cap \mathrm{Ker}(V)$, and

look instead for occurences of distinct-coordinates values of ψ in A^t. The canonical setting of study is that of the single translation-invariant equation $y_1 + y_3 = 2y_2$, which detects 3-term arithmetic progressions, themselves parametrized by the system of forms

$$\psi(x_1, x_2) = (x_1, x_1 + x_2, x_1 + 2x_2).$$

It is then a well-known result of Green [30] that every subset of $\mathcal{P}_N$ of positive density contains a non-trivial three-term arithmetic progression; and the extension of this result to progressions of any length is the celebrated Green-Tao theorem [36]. Green's argument [30] actually allowed for densities as low as $(\log\log\log\log N)^{-1/2+o(1)}$, and Helfgott and de Roton [50] later obtained a remarkable quantitative strenghtening of this result.

THEOREM 1.1 (Helfgott, de Roton). *Suppose that A is a subset of $\mathcal{P}_N$ of density at least*[1]

$$(\log\log N)^{-1/3+o(1)}.$$

Then there exists a non-trivial three-term arithmetic progression in A.

Naslund [65] further improved the lowest admissible density to $(\log\log N)^{-1+o(1)}$. It should be noted that these transference arguments preserve, up to a logarithm, the exponent in the best known bounds for Roth's theorem by Sanders [81], on which they rely: indeed Sanders established that three-term arithmetic progressions may be found in any subset of $[N]$ of density at least $(\log N)^{-1+o(1)}$.

In the context of counting linear patterns in primes [39], Green and Tao introduced the notion of *Cauchy-Schwarz complexity*[2] (abbreviated as complexity in the

[1] Throughout this introduction, we write $(\log_k N)^{o(1)}$ for unspecified factors of the form $C(\log_{k+1} N)^C$ with $C > 0$, where $\log_k$ is the k-th iterated logarithm.

[2] A more subtle notion of complexity, called *true complexity*, was later developed by Gowers and Wolf [23]. However it does not seem, at present, to cover the setting of unbounded prime-counting functions.

following) for systems of integer linear forms. Precisely, we say that a system of t distinct linear forms $(\psi_1, \dots, \psi_t)$ has complexity at most s when, for every $i \in [t]$, it is possible to partition the set of forms $\{\psi_j, j \neq i\}$ into at most $s+1$ sets, such that ψ_i does not belong to the linear span of any of those sets. The condition of finite complexity is then equivalent to requiring that no two forms of the system be linearly dependent. By extension, we define the complexity of a matrix V to be that of any parametrization $\psi : \mathbb{Z}^d \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}(V)$, this property being independent of the choice of ψ.

Systems of complexity at most one may be analyzed by methods of classical Fourier analysis, whereas cases of higher complexities require much more involved techniques [20, 35]. We focus on the case of complexity one here, for it is possible to derive strong quantitative bounds in that setting, and for it may provide insight on how to quantify results of higher complexity. On the qualitative side, it is known that a translation-invariant system of equations $V\mathbf{y} = 0$ of finite complexity is non-trivially solvable in any subset of the primes of positive upper density: this follows from the Green-Tao theorem [36] on arithmetic progressions in the primes, by an elementary argument discussed in Section 10. Our main finding is that, in the case of complexity one, quantitative bounds of the quality of Helfgott and de Roton's may be achieved.

THEOREM 1.2. *Let $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ be a translation-invariant matrix of rank r and complexity one. There exists a positive constant C depending at most on r, t, V such that, if A is a subset of $\mathcal{P}_N$ of density at least*

$$C(\log\log N)^{-1/25t},$$

there exists $\mathbf{y} \in A^t$ with distinct coordinates such that $V\mathbf{y} = 0$.

Our argument also preserves the aforementioned feature of Naslund's refinement of the Helfgott-de Roton transference principle: in the complexity one regime, it converts logarithmic density bounds $(\log N)^{-\gamma}$ for Szemerédi-type theorems in the

integers, to doubly logarithmic bounds $(\log\log N)^{-\gamma+\varepsilon}$ for Szemerédi-type theorems in the primes. We mention however that Theorem 1.2 is surpassed, in certain special cases, by results in the integers. Indeed, an important result of Schoen and Shkredov [90] states that any single translation-invariant equation in a least 6 variables is non-trivially solvable in any subset of $[N]$ of density $e^{-(\log N)^{1/6-o(1)}}$, and hence in $\mathcal{P}_N$, however it is not clear whether or how that result extends to the case of several equations. Furthermore, in certain "degenerate" cases where the $r \times t$ matrix V may be subdivided into translation-invariant $r \times t_i$ submatrices, the system of equations may even be solvable at densities N^{-c}: we refer to the work of Shapira [92], generalizing that of Ruzsa [76], for precise statements.

To motivate Theorem 1.2, we now give some illustrative examples of systems of complexity one. First, any single translation-invariant equation has complexity one, although in that case a simple modification of the argument of Helfgott and de Roton [50] yields Theorem 1.2. A more representative example of a system of complexity one is that of "d points and their midpoints", corresponding to the set of equations $(y_{ii} + y_{jj} = 2y_{ij})_{1\leqslant i<j\leqslant d}$, whose solutions over $\mathbb{Q}$ are parametrized, with some multiplicity, by[3] $\psi(x) = (x_0 + x_i + x_j)_{1\leqslant i\leqslant j\leqslant d}$. It can be arduous in general to determine whether a system of equations has complexity one: Vinuesa [105] has determined, by an elaborate combinatorial argument, that the system of translation-invariant equations corresponding to magic $n \times n$ squares has complexity one for $n \geqslant 4$. Besides specific examples, there also exists a strong set of conditions on the matrix V designed by Roth [70], which allows for a Fourier analysis of translation-invariant equations; in particular, these conditions are satisfied for matrices $V \in \mathcal{M}_{r\times(2r+1)}(\mathbb{Z})$ containing only invertible $r \times r$ submatrices, and such matrices have complexity one. Roth's conditions have received further attention in work of Liu, Spencer and Zhao [61, 62] and in Section 9, we compare those

[3] This system is the linear part of Example 4 from [39, Section 1], composed with a certain surjection.

conditions to the assumption of complexity one, showing in particular that a slight strengthening of the former implies the latter.

Next, we discuss the principal ideas behind the proof of Theorem 1.2. The main structure of our argument follows the ubiquitous transference principle [30, 36], by which one lifts a dense subset of the primes to a dense subset of the integers. More precisely, we initially follow the transference strategy of Helfgott and de Roton [50], incorporating also Naslund's [65] sharper estimates. Denoting by λ_A the renormalized indicator function of a dense subset A of the primes, we therefore compare the average of λ_A over ψ-patterns to that of a smoothed version λ'_A of itself, which behaves as a dense subset of the integers of almost the same density. As usual, there is a little technical subtelty in the form of the W-trick, by which we consider, instead of the set A, its intersection with an arithmetic progression of modulus $W = \prod_{p \leqslant \omega} p$. A critical feature of Helfgott and de Roton's argument [50] is then that it requires a modulus $\omega \sim c \log N$.

At this point we invoke a beautiful recent result of Shao [91], who improved on a first result of Dousse [15], and generalized the logarithmic bounds of Bourgain [5] for Roth's theorem to a model system of complexity one. More precisely, Shao [91] investigated the system $\psi(x) = (x_0 + x_i + x_j)_{1 \leqslant i \leqslant j \leqslant d}$, and proved that a set A of density $(\log N)^{-1/6d(d+1)+o(1)}$ in $[N]$ contains a non-trivial configuration $\psi(x) \in A^{d(d+1)/2}$. As envisioned by Shao [91, p. 2], his argument naturally extends to general systems of complexity one, at the cost of adressing certain technical complications. The first, and simplest step of our proof is therefore to formally derive this extension, while also keeping track of the number of pattern occurences. Considering λ'_A as a dense set of integers, this extension then shows that λ'_A has a large pattern count.

Provided that we could prove that the difference of pattern counts for λ_A and λ'_A is small, this would be enough to conclude that the original set A contains many ψ-configurations. However, while the count of three-term progressions investigated by Helfgott and de Roton [50] has a simple Fourier expression, which can be

controlled by restriction estimates for primes [34], such is not the case in general for systems of complexity one. To address this issue, we bound the difference of pattern counts via the generalized Von Neumann theorem of Green and Tao [39], which in the complexity-one setting asserts that, given functions $f_1, \dots, f_t$ on $\mathbb{Z}_{N'}$ with $N' \sim CN$ majorized by a pseudorandom weight (a notion whose meaning shall be clear shortly), we have

$$\left|\mathbb{E}_{n\in\mathbb{Z}_{N'}^d} f_1(\psi_1(n)) \dots f_t(\psi_t(n))\right| \leqslant \|f_i\|_{U^2} + o(1) \tag{1.1}$$

as $N \to \infty$. Properly quantified, the method of Green and Tao [36, 39] produces a $o(1)$ term of size $(\log N)^{-c}$ in the above, however it requires a small modulus $\omega \sim c \log\log N$, which is too expensive to apply the efficient transference estimates of Helgott and de Roton [50].

To majorize prime-counting functions associated to W-tricked primes, Green and Tao use a weight $\nu : \mathbb{Z}_M \to \mathbb{R}^+$ constructed from a smoothly truncated convolution of the Möbius function, which was first considered by Goldston, Pintz and Yildirim [19]. The $o(1)$-term arising in (1.1) then depends on the level of pseudorandomness of this weight, and the key estimate we establish towards this is the asymptotic

$$\mathbb{E}_{n\in\mathbb{Z}_{N'}^d} \nu(\theta_1(n)) \dots \nu(\theta_t(n)) = 1 + O_{d,t,\theta}\left(\frac{1}{(\log N)^{1-o(1)}}\right),$$

valid for every affine system $\theta : \mathbb{Z}_{N'}^d \to \mathbb{Z}_{N'}^t$ of finite complexity and bounded linear part, and for a large modulus $\omega \sim c \log N$. This corresponds to the "linear forms condition" in [36, 39], while we do not need the harder-to-quantify "correlation condition" from there in our simpler setting. Equipped with this estimate, we verify that the functions λ_A and λ'_A used by Helfgott and de Roton are majorized by averaged variants of ν, and we finally apply (1.1) to bound the difference of pattern counts.

Remarks. Very recently, and while we were writing this article, Conlon, Fox and Zhao have completed an exposition of the Green-Tao theorem [8], in which they also revisited Green and Tao's computations on correlations of GPY weights under the assumption of finite complexity. Their number-theoretic computations [8, Section 9] turn out to be very similar to ours from Section 5, although our argument optimizes certain parameters further.

Acknowledgements. We are grateful to our adviser Régis de la Bretèche for valuable advice on writing. We also wish to thank our friends Crystel Bujold, Dimitri Dias, Oleksiy Klurman, Marzieh Mehdizad for helpful discussions on many topics of number theory. We would further like to thank Pablo Candela, Harald Helfgott, Neil Lyall, Eric Naslund, Hans Parshall and Fernando Shao for interesting discussions on problems related to this paper.

2. Overview

In this section we explain the organization of this paper, and we outline in more detail the structure of our argument, previously sketched in the introduction.

The preliminaries to our argument are contained in Sections 3 and 4. The little notation we need is introduced in Section 3, while Section 4 is there to gather (almost) all arguments of a linear algebraic nature needed in the article.

As is traditional in additive combinatorics, we then delegate to appendices material which is either relatively standard or not fully relevant to the main text. Thus, in Section 8, we derive the aforementioned extension of Shao's [91] result, and in Section 10 we derive, for the comfort of the reader, several results on translation-invariant equations which are known to follow from the literature. In Section 9, we study the notion of complexity one in more detail. That Appendix is not formally needed for the proof of Theorem 1.2, however it sheds light on the class of systems to which it applies.

The bulk of our proof of Theorem 1.2 is therefore contained in Sections 5–7. In Section 5, we carry out the computation of correlations of the GPY weights

$$\Lambda_{\chi,R,W}(n) = \Big(\frac{\phi(W)}{W}\log R\Big)\Big(\sum_{d|Wn+b}\mu(d)\chi\Big(\frac{\log d}{\log R}\Big)\Big)^2,$$

where $W = \prod_{p\leqslant\omega} p$ and χ is a certain smooth cutoff function. We follow Green and Tao's original computation [**39**, Appendix D], but we analyze the local Eulor factors involved in more detail, in order to allow for a large modulus $\omega = c\log N$. In Section 6, we construct a pseudorandom weight on ν over $\mathbb{Z}_M$ out of $\Lambda_{\chi,R,W} : \mathbb{Z} \to \mathbb{R}^+$ for a larger scale $M \sim CN$, taking care to preserve quantitative error terms. We also state a quantitative version of Green and Tao's generalized Von Neumann theorem [**39**, Appendix C]. In Section 7, we prove Theorem 1.2, by first lifting the problem to the integers via the transference principle of Helfgott-de Roton [**50**] and the quantitative generalized Von Neumann theorem obtained earlier, and by then applying the extension of Shao's result derived in Section 8.

3. Notation

We have attempted to respect most current conventions of notation in additive combinatorics [**27**] throughout, and therefore we keep this section to the bare minimum.

Given an integer N, we write $[N] = \{1,\dots,N\}$. Given reals $x < y$, we also write $[x,y]_{\mathbb{Z}} = \mathbb{Z}\cap[x,y]$, and we let $\mathcal{P}$ denote the set of all primes. Given a property $\mathbf{P}$, we write $1(\mathbf{P})$ for the boolean which equals 1 when $\mathbf{P}$ is true, and 0 otherwise. When X is a set and $\mathbf{P}_x$ is a property depending on a variable $x \in X$, we write

$$\mathbb{P}_{x\in X}(\mathbf{P}_x) = |X|^{-1}\#\{x\in X : \mathbf{P}_x\}.$$

Given a function f on X, we also write $\mathbb{E}_X f = \mathbb{E}_{x\in X} f(x) = |X|^{-1}\sum_{x\in X} f(x)$, or simply $\mathbb{E}f$ when the set of averaging is clear from the context.

We make occasional use of Landau's o, O-notation and of Vinogradov's asymptotic notations $f \ll g$, $f \gg g$, $f \asymp g$. As is common in additive combinatorics, we also let c and C denote positive constants whose value may change at each occurence, and which are typically taken to be respectively very small or very large. Unless otherwise stated, all implicit and explicit constants we introduce are absolute: they do not depend on surrounding parameters.

Finally, we use several local conventions on notation, and therefore we advise the reader to pay close attention to the preamble of each section.

4. Linear algebra preliminaries

In this section, we discuss the notion of complexity of systems of linear forms, following the very transparent exposition by Green and Tao in [39, Sections 1 and 4], and by Tao in [98]. We also consider the simple problems of parametrizing the kernel of a matrix corresponding to a system of equations, and of defining an analog notion of complexity for such a matrix.

We consider an integral domain $\mathbb{A}$, together with its field of fractions $\mathbb{K}$; in our article we only ever consider $\mathbb{A} = \mathbb{Z}$ or $\mathbb{A} = \mathbb{Z}_M$ with M prime. A linear form over the free module $\mathbb{A}^d$ naturally induces one over $\mathbb{K}^d$, and accordingly all the linear algebra notions are considered over $\mathbb{K}$. This is somewhat overly formal, however it allows us to define certain notions for linear forms over $\mathbb{Z}$ and $\mathbb{Z}_M$ at once. Note that throughout this article, we consider systems of linear forms $\psi : \mathbb{A}^d \to \mathbb{A}^t$ as formal triples (ψ, d, t) to avoid repeatedly introducing dimension parameters d, t.

DEFINITION 4.1 (Complexity). *Consider a system of linear forms $\psi = (\psi_1, \dots, \psi_t) : \mathbb{A}^d \to \mathbb{A}^t$. For $i \in [t]$, the complexity of ψ at i is the minimal integer $s \geqslant 0$ for which there exists a partition $[t] \smallsetminus \{i\} = X_1 \sqcup \cdots \sqcup X_{s+1}$ into non-empty sets such that $\psi_i \notin \langle \psi_j : j \in X_k \rangle$ for all $k \in [s+1]$, when such an integer exists*[4]*. Otherwise we set*

[4] In the special (and unimportant) case where $t = 1$, we set the complexity at $i = 1$ to 0.

the complexity at i to ∞. The complexity of ψ is the maximum of the complexities of ψ at i over all $i \in [t]$.

We also recall the following important observation from [39, Section 1].

LEMMA 4.2. *A system of linear forms $\psi = (\psi_1, \dots, \psi_t) : \mathbb{A}^d \to \mathbb{A}^t$ has finite complexity if and only if no two forms ψ_i, ψ_j with $i \neq j$ are linearly dependent.*

We next recall the standard notion of normal form, and to do so we introduce a slightly non-standard piece of terminology. We say that a linear form $\theta(x_1, \dots, x_d) = a_1x_1 + \cdots + a_dx_d$ *depends* on the variable x_k when $a_k \neq 0$; we do not mean this in an exclusive sense so that the form may also depend on other variables. While that definition may seem mathematically akward, it corresponds to the intuitive way to think about explicit system of forms.

DEFINITION 4.3 (Normal form). *A system of linear forms $\psi = (\psi_1, \dots, \psi_t) : \mathbb{A}^d \to \mathbb{A}^t$ is in* exact *s-normal form at $i \in [t]$ when there exists a set of indices $J_i \subset [d]$ such that $|J_i| = s + 1$ and*

(i) *$\psi_i(x_1, \dots, x_d)$ depends on all variables $x_k, k \in J_i$,*

(ii) *for all $j \neq i$, $\psi_j(x_1, \dots, x_d)$ does not depend on all variables $x_k, k \in J_i$.*

We say that ψ is in s-normal form when it is in exact s_i-normal form with $s_i \leqslant s$ at every $i \in [t]$.

As explained in [39, Section 4], a system ψ in exact s-normal form at i has complexity at most s at i, and conversely one may always put a system of complexity s in s-normal form, up to adding a certain number of "dummy" variables.

PROPOSITION 4.4 (Normal extension). *A system of linear forms $\psi : \mathbb{Z}^d \to \mathbb{Z}^t$ of complexity s admits an s-normal extension $\psi' : \mathbb{Z}^{d+e} \to \mathbb{Z}^t$ of the form $\psi'(x, y) = \psi(x + \varphi(y))$, where $\varphi : \mathbb{Z}^e \to \mathbb{Z}^d$ is a linear form.*

We will also have the occasion to consider systems of affine-linear forms, often abbreviated as "affine systems" throughout the article. Consistently with [39], we

write an affine system ψ as $\psi = \psi(0) + \dot{\psi}$, where $\dot{\psi}$ is the linear part of ψ, and we extend previous definitions by declaring ψ to be of complexity s or in s-normal form when its linear part is. We also need to consider reductions of forms modulo a large prime M later on, in which case we need to keep track of the size of the coefficients of the forms involved.

DEFINITION 4.5 (Form and matrix norms). *Suppose that $\psi = (\psi_1, \dots, \psi_t) : \mathbb{A}^d \to \mathbb{A}^t$ is an affine system, and write $\psi_i(x_1, \dots, x_d) = a_{i1}x_1 + \cdots + a_{id}x_d + b_i$ for every $i \in [t]$. When $\mathbb{A} = \mathbb{Z}$ and $M \geqslant 1$, we define*

$$\|\psi\|_M = \sum_{i\in[t]} \sum_{j\in[d]} |a_{ij}| + \sum_{i\in[t]} (|b_i|/M),$$

and we simply write $\|\psi\|$ when all b_i are zero. When $\mathbb{A} = \mathbb{Z}_M$, we define

$$\|\psi\| = \sum_{i\in[t]} \sum_{j\in[d]} \|a_{ij}\|_{\mathbb{T}_M} + \sum_{i\in[t]} \|b_i/M\|_{\mathbb{T}}$$

where $\|\cdot\|_{\mathbb{T}_L} = d(\cdot, L\mathbb{Z})$. Finally, for a matrix $V = [\lambda_{ij}] \in \mathcal{M}_{r\times t}(\mathbb{Z})$, we write

$$\|V\| = \sum_{i,j} |\lambda_{ij}|.$$

We now return to our main topic of interest, that is, translation-invariant equations in the integers. As for systems of forms, we consider matrices $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ as formal triples (V, r, t).

DEFINITION 4.6. *We say that $V = [a_{ij}] \in \mathcal{M}_{r\times t}(\mathbb{Z})$ is translation-invariant when*

$$a_{i1} + \cdots + a_{it} = 0 \quad \forall i \in [r].$$

Given a matrix $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ corresponding to a system of equations $V\mathbf{y} = 0$, we now define the complexity of V at an indice $i \in [t]$, and its global complexity, to be that of any system of linear forms $\psi : \mathbb{Q}^d \twoheadrightarrow \mathrm{Ker}(V)$. The following proposition ensures that such a definition does not depend on the choice of parametrization ψ.

PROPOSITION 4.7 (Matrix complexity criterion). *Consider a matrix $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ with lines $L_1, \dots, L_r$ and $t \geqslant 2$, and a system of linear forms $\psi : \mathbb{Q}^d \twoheadrightarrow \mathrm{Ker}(V)$. Then ψ has complexity at most s_0 at i if and only if there exists $0 \leqslant s \leqslant s_0$ and a partition $[t] \smallsetminus \{i\} = X_1 \sqcup \cdots \sqcup X_{s+1}$ into non-empty sets such that, for every $k \in [s+1]$,*

$$\left(e_i + \textstyle\sum_{j\in X_k} \mathbb{Q}e_j\right) \cap \langle {}^{\mathrm{t}}L_1, \dots, {}^{\mathrm{t}}L_r\rangle = \varnothing,$$

where $(e_i)_{1\leqslant i\leqslant t}$ is the canonical basis of $\mathbb{Q}^t$.

PROOF. Consider $i \in [t]$ and a partition $[t] \smallsetminus \{i\} = X_1 \sqcup \dots X_{s+1}$ into non-empty sets. For any $k \in [s+1]$ and $\lambda \in \mathbb{Q}^{X_k}$, we have an equivalence

$$\begin{aligned} & \psi_i + \textstyle\sum_{j\in X_k} \lambda_j \psi_j = 0 \\ \Leftrightarrow \quad & x_i + \textstyle\sum_{j\in X_k} \lambda_j x_j = 0 \text{ for all } x \in \mathrm{Ker}(V) \\ \Leftrightarrow \quad & e_i + \textstyle\sum_{j\in X_k} \lambda_j e_j \in \mathrm{Ker}(V)^\perp. \end{aligned}$$

Furthermore, by orthogonality in $\mathbb{Q}^t$,

$$\mathrm{Ker}(V)^\perp = \left(\langle {}^{\mathrm{t}}L_1, \dots, {}^{\mathrm{t}}L_t\rangle^\perp\right)^\perp = \langle {}^{\mathrm{t}}L_1, \dots, {}^{\mathrm{t}}L_r\rangle.$$

Therefore $\psi_i \in \langle \psi_j, j \in X_k\rangle$ if and only if there exists $\lambda \in \mathbb{Q}^{X_k}$ such that $e_i + \sum_j \lambda_j e_j \in \langle {}^{\mathrm{t}}L_1, \dots, {}^{\mathrm{t}}L_r\rangle$. The proposition follows by considering the contrapositive. □

We shall have the occasion to work with two standard types of parametrizations for the integer kernel of a translation-invariant matrix. The first is the usual normal form, which is useful when working with primes, while the second has an added shift variable, which is useful for the regularity computations of Section 8. In both cases, it is critical to work with a base parametrization ψ in normal form, in order to bound averages over patterns $(\psi_1(n), \dots, \psi_t(n))$ by a certain Gowers norm (see Propositions 6.4 and 8.10 below).

PROPOSITION 4.8 (Kernel parametrization). *Suppose that $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ is a translation-invariant matrix of rank r and complexity at most s. Then there exists a linear surjection*

$$\psi : \mathbb{Z}^d \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}(V)$$

in s-normal form. An alternate linear surjection is then given by

$$\varphi : \mathbb{Z}^{d+1} \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}(V),$$

where φ is defined by $\varphi_i(x_0, x) = x_0 + \psi_i(x)$ for every $i \in [t]$ and $(x_0, x) \in \mathbb{Z} \times \mathbb{Z}^d$.

PROOF. The set $\mathbb{Z}^t \cap \mathrm{Ker}(V)$ is a lattice which is easily seen to be of rank $t - r$ (e.g. by first solving $V\mathbf{y} = 0$ over $\mathbb{Q}$, then clearing denominators), so that there exists a linear isomorphism $\psi : \mathbb{Z}^{t-r} \xrightarrow{\sim} \mathbb{Z}^t \cap \mathrm{Ker}(V)$ of complexity at most s. Since extensions in the sense of Proposition 4.4 preserve the image of a form, we may choose an alternate linear parametrization $\psi' : \mathbb{Z}^d \xrightarrow{\sim} \mathbb{Z}^t \cap \mathrm{Ker}(V)$ in s-normal form for a certain $d \geqslant t - r$.

Since the matrix V is translation-invariant, we have $V\mathbf{1} = 0$, where $\mathbf{1} = (1, \dots, 1)$. Therefore we may define another surjection $\varphi : \mathbb{Z} \times \mathbb{Z}^d \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}(V)$ by $\varphi(x_0, x) = x_0\mathbf{1} + \psi'(x)$. □

Note that a system of linear forms $\psi : \mathbb{Z}^d \to \mathbb{Z}^t$ in 1-normal form is, at every position $i \in [t]$, either in exact 0-normal form or in exact 1-normal form. In practice we can always eliminate the first possibility, and while not of fundamental importance, this fact allows us to simplify our argument in some places.

PROPOSITION 4.9. *Suppose that $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ is a matrix of complexity one with no zero columns and $t \geqslant 3$, and $\psi : \mathbb{Z}^d \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}(V)$ is a system of linear forms in 1-normal form. Then ψ is in exact 1-normal form at every $i \in [t]$.*

PROOF. This follows from the complexity-zero criterion of Proposition 9.3, and the fact that s-normality at i implies complexity at most s at i for any $i \in [t]$. □

One last simple fact we require about (translation-invariant) systems of equations is a bound on the number of integer solutions with two equal coordinates in a box.

LEMMA 4.10 (Number of degenerate solutions). *Suppose that $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ has rank r and finite complexity, and let i, j be two distinct indices in $[t]$. Then*

$$\#\{y \in [-N, N]^t_{\mathbb{Z}} : Vy = 0 \text{ and } y_i = y_j\} \ll_V N^{t-r-1}.$$

PROOF. Consider the hyperplane $H = \{y \in \mathbb{Q}^t : y_i = y_j\}$. The subspace $\mathrm{Ker}(V)\cap H$ of $\mathbb{Q}^t$ has dimension less than $t-r-1$, since $\mathrm{Ker}(V)$ is not contained in H: indeed if this were the case, there would exist a parametrization $\psi : \mathbb{Z}^d \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}(V)$ with $\psi_i = \psi_j$, contradicting the assumption of finite complexity. The bound then follows by simple linear algebraic considerations. □

Finally, we collect together some facts about the preservation of certain properties of affine systems under the operations of reduction modulo M or lifting from $\mathbb{Z}_M$ to $\mathbb{Z}$. We omit the proofs, which are accessible by simple linear algebra.

FACT 4.11. *Suppose that $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ is a translation-invariant matrix of rank r and $\psi : \mathbb{Z}^d \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}_{\mathbb{Q}}(V)$ is a system of linear forms in exact s_i-normal form over $\mathbb{Z}$ at every $i \in [t]$. Provided that $M > \max(t!\|\psi\|^t, r!\|V\|^r)$, ψ reduces modulo M to a system of linear forms $\theta : \mathbb{Z}_M^d \twoheadrightarrow \mathrm{Ker}_{\mathbb{Z}_M}(V)$ is in exact s_i-normal form over $\mathbb{Z}_M$ at every $i \in [t]$, and such that $\|\theta\| = \|\psi\|$.*

FACT 4.12. *Suppose that $\theta : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ is an affine system of finite complexity over $\mathbb{Z}_M$, and $M > 2\|\dot\theta\|$. Then θ is the reduction modulo M of an affine system $\psi : \mathbb{Z}^d \to \mathbb{Z}^t$ of finite complexity over $\mathbb{Z}$ and such that $\|\psi\|_M = \|\theta\|$, $\|\dot\psi\| = \|\dot\theta\|$.*

5. Correlations of GPY weights

The aim of this section is to construct efficient pseudorandom weights over $\mathbb{Z}$ majorizing the measure associated to W-tricked primes. The weight we consider (see Definition 5.3 below) is a truncated divisor sum whose correlations were first

investigated by Goldston, Pintz and Yildirim [19] in the context of small gaps between primes. Green and Tao [36, 39] further investigated its pseudorandom behavior, through more sophisticated correlation computations, and this weight is by now a standard tool, e.g. in the context of detecting polynomial patterns in primes [59, 101, 102].

Throughout this section, we consider an integer N larger than some absolute constant, and we let $\omega \geqslant 1$ be a parameter. We also let $W = \prod_{p \leqslant \omega} p$ and we fix an integer b such that $(b, W) = 1$. It is then useful to have a notation for the normalized indicator function of W-tricked primes.

DEFINITION 5.1 (Measure of W-tricked primes). *We let*

$$\lambda_{b,W}(n) = \frac{\phi(W)}{W}(\log N) \cdot 1(n \in [N] \text{ and } b + Wn \in \mathcal{P}).$$

Our goal is thus to construct a weight function over $\mathbb{Z}$ majorizing $\lambda_{b,W}$, and satisfying strong pseudorandomness asymptotics. Note that $o(1)$ terms throughout this article are to be understood as $N \to \infty$, and do not depend on any dimension or any affine system involved.

PROPOSITION 5.2 (Pseudorandom majorant over $\mathbb{Z}$). *Let $D \geqslant 1$ be a parameter. There exists a constant C_D such that the following holds. For $N \geqslant C_D$ and $\omega = c_0 \log N$, there exists $\nu : \mathbb{Z} \to \mathbb{R}^+$ such that, for every $\varepsilon > 0$,*

$$0 \leqslant \lambda_{b,W} \ll_D \nu \ll_\varepsilon N^\varepsilon$$

and, for any $P \geqslant N^{c_1}$ and any affine system $\psi : \mathbb{Z}^d \to \mathbb{Z}^t$ of finite complexity and such that $d, t, \|\dot{\psi}\| \leqslant D$,

$$\mathbb{E}_{n \in [P]^d} \nu\big[\psi_1(n)\big] \dots \nu\big[\psi_t(n)\big] = 1 + O_D\left(\frac{1}{(\log N)^{1-o(1)}}\right). \tag{5.1}$$

Note that simply applying [39, Theorem D.3] would be insufficient for our purpose, since the error there is $e^{O(\sqrt{\omega})}(\log N)^{-1/20}$ and therefore it is non-trivial

only for $\omega \leqslant c(\log\log N)^2$, thus rendering the methods of Helfgott and de Roton [50] unapplicable. The argument of [36] also requires a modulus $\omega \leqslant c\log\log N$. Our construction follows closely that in [39, Appendix D], however with one important difference: we make a stronger assumption of finite complexity on the system of linear forms, and under this assumption we obtain improved estimates on the Euler products involved. We also remark that for the purpose of proving Theorem 1.2, any error term of the form $(\log N)^{-c}$ in (5.1) would suffice, however we take the opportunity here to determine the highest level of pseudorandomness attainable from Green and Tao's approach.

We let $\chi \in C^\infty(\mathbb{R})$ denote a certain positive function with $\chi(0) = 1$ and support in $[-1, 1]$, and we consider an additional parameter $1 \leqslant R \leqslant N$. Our main object of study in this section is the following weight function.

DEFINITION 5.3 (GPY weight). *We let* $h_{R,W} = \frac{\phi(W)}{W}\log R$ *and*

$$\Lambda_{\chi,R,W}(n) = h_{R,W}\bigg(\sum_{m|Wn+b}\mu(m)\chi\Big(\frac{\log m}{\log R}\Big)\bigg)^2.$$

The pseudorandom weight we seek will turn out to be a scalar multiple of the above function: we defer the precise choice of normalization until the end of the proof of Proposition 5.2.

LEMMA 5.4. *When* $\omega = c_0\log N$ *and* $R = N^\eta$ *with* $0 < \eta \leqslant c_0/2$, *we have*

$$0 \leqslant \lambda_{b,W} \ll_\eta \Lambda_{\chi,R,W} \ll_\varepsilon N^\varepsilon$$

for every $\varepsilon > 0$.

PROOF. If $\lambda_{b,W}(n)$ is non-zero, $Wn + b$ is a prime of size at least $W > N^{c_0/2}$, for N large enough. Therefore any non-trivial divisor of $Wn+b$ has size larger than R, so that $\Lambda_{\chi,R,W}(n) = \frac{\phi(W)}{W}(\log R)\chi(0) \leqslant \eta^{-1}\lambda_{b,W}(n)$. The last inequality follows from standard bounds on the divisor function [103]. □

We now say more on the choice of cutoff function χ. We start by picking a smooth positive function $F \in C_c^\infty(\mathbb{R})$ with $F(0) = 1$ and support in $[-1, 1]$, and such that[5] $\widehat{F}(\xi) \ll e^{-c|\xi|^{1/2}}$ uniformly in $\xi \in \mathbb{R}$; there are various well-known constructions of such functions [28, 55]. We then define $\chi(x) = e^x F(x) \in C_c^\infty(\mathbb{R})$, so that by Fourier inversion we may write

$$\chi(x) = \int_{-\infty}^{\infty} \varphi(\xi) e^{-(1+i\xi)x} \mathrm{d}\xi \qquad (x \in \mathbb{R}), \tag{5.2}$$

where φ is a certain integrable function satisfying the decay estimate[6]

$$\varphi(\xi) \ll e^{-c|\xi|^{1/2}} \qquad (\xi \in \mathbb{R}). \tag{5.3}$$

We now begin the proof of Proposition 5.2. We fix $D \geqslant 1$ and $\omega = c_0 \log N$, so that we may assume that ω is larger than any fixed constant depending on D. We then consider a system of affine-linear forms $\psi : \mathbb{Z}^d \to \mathbb{Z}^t$ of finite complexity such that $d, t, \|\dot{\psi}\| \leqslant D$. We let further implicit constants and explicit unsuscripted constants c, C depend on $d, t, \|\dot{\psi}\|$, while subscripted constants $c_0, c_1, \dots$ are absolute.

The first step of the proof is to unfold divisor sums in the correlation of divisor sums, and it is useful in this regard to introduce the notation $\Omega = [t] \times [2]$. Note also that the prime in $\sum'$ means that the summation is restricted to square-free numbers. The following constitutes the beginning of the proof of [39, Theorem D.3], which we do not reproduce.

PROPOSITION 5.5 (Unfolding sums). *Given $(m_{ij}) \in \mathbb{N}^\Omega$, write $m_i = [m_{i1}, m_{i2}]$ and*

$$\alpha(m_1, \dots, m_t) = \mathbb{P}_{n \in \mathbb{Z}_m^d} \Big(m_i | W\psi_i(n) + b \quad \forall i \in [t] \Big).$$

[5] Here $\widehat{F}(\xi) = \int_{\mathbb{R}} F(x) e(-\xi x) \mathrm{d}x$.

[6] Using a weaker decay $\ll (1 + |\xi|)^{-A}$ instead would yield a slightly weaker error term $(\log N)^{-1+\varepsilon}$ in Proposition 5.2.

Let also $P \geqslant 1$. Then

$$\begin{aligned} &h_{R,W}^{-t} \sum_{n\in[P]^d} \Lambda_{\chi,R,W}\big[\psi_1(n)\big]\dots\Lambda_{\chi,R,W}\big[\psi_t(n)\big] \\ =&P^d \cdot \sum_{(m_{ij})\in\mathbb{N}^\Omega}{}' \alpha(m_1,\dots,m_t) \prod_{(i,j)\in\Omega} \mu(m_{ij})\chi\Big(\frac{\log m_{ij}}{\log R}\Big) + O(R^{2|\Omega|}P^{d-1}) \end{aligned}$$

Before proceeding further, we analyze the function α appearing in Proposition 5.5. By the Chinese Remainder theorem, $\alpha(m_1,\dots,m_t)$ is multiplicative in the variables m_{ij}, keeping in mind that $m_i = [m_{i1}, m_{i2}]$. Writing $m_{ij} = p^{r_{ij}}$, $r_i = \max(r_{i1}, r_{i2})$, and $B = \{(i,j) \in \Omega : r_{ij} = 1\}$, we have $r_i = 1$ if and only if $r_{ij} = 1$ for some $j \in [2]$, that is, if and only if the slice B_i of B at i is non-empty. Therefore

$$\alpha(p^{r_1},\dots,p^{r_t}) = \mathbb{P}_{n\in\mathbb{Z}_p^d}\big(p|W\psi_i(n)+b \quad \forall i : B_i \neq \varnothing\big) =: \alpha(p,B). \tag{5.4}$$

Motivated by this, we say that a non-empty set $B \subset \Omega$ is *vertical* when, for some $i \in [t]$, we have $B \subset \{i\} \times [2]$. We now estimate the size of the factors $\alpha(p,B)$.

PROPOSITION 5.6 (Local probabilities). *For $B \neq \varnothing$, we have*

$$\alpha(p,B) = \begin{cases} 0 & \text{if } p \leqslant \omega \\ p^{-1} & \text{if } p > \omega \text{ and } B \text{ is vertical} \\ O(p^{-2}) & \text{if } p > \omega \text{ and } B \text{ is not vertical} \end{cases}$$

PROOF. Recall that $\alpha(p,B)$ is defined by (5.4). When $p \leqslant \omega$, we have $p|W$ and $(b,W) = 1$, therefore p does not divide any value $W\psi_i(n) + b$ and $\alpha(p,B) = 0$. When $p > \omega > \|\dot\psi\|$, we have $p \nmid W$ and $W\dot\psi_i \neq 0$ in $\mathbb{Z}_p$ for every $i \in [t]$. When B is vertical, there is only one i such that B_i is non-empty and therefore $\alpha(p,B) = p^{-1}$, since hyperplanes of $\mathbb{Z}_p^d$ have size p^{d-1}. When B is not vertical, there are at least two indices i,j such that $B_i, B_j \neq \varnothing$. Since $p > \omega > 2\|\dot\psi\|^2$, the linear forms $\dot\psi_i$ and $\dot\psi_j$ are linearly independent over $\mathbb{Z}_p$, therefore $\alpha(p,B) \leqslant p^{-2}$ since $(d-2)$-flats of $\mathbb{Z}_p^d$ have size p^{d-2}. □

For reasons that shall be clear in a moment, we define the following Euler factor.

DEFINITION 5.7 (Euler factor). *Let $\xi \in \mathbb{R}^\Omega$ and $z_{ij} = (1 + i\xi_{ij})/\log R$. We let*

$$E_{p,\xi} = \sum_{B \subset \Omega} (-1)^{|B|} \alpha(p, B) p^{-\sum_{(i,j)\in B} z_{ij}}. \tag{5.5}$$

The local estimates of Proposition 5.6 and the fact that $\mathrm{Re}(z_{ij}) > 0$ ensure the absolute convergence of the product $\prod_p E_{p,\xi}$. We now return to the unfolded sum in Proposition 5.5, in which we proceed to replace the weights χ by truncations of their Fourier expression.

PROPOSITION 5.8 (Unfolding integrals). *Writing $m_i = [m_{i1}, m_{i2}]$, we have, for any $L \geqslant 1$,*

$$\sum_{(m_{ij})\in\mathbb{N}^\Omega}{}' \alpha(m_1, \dots, m_t) \prod_{(i,j)\in\Omega} \mu(m_{ij}) \chi\Big(\frac{\log m_{ij}}{\log R}\Big) \tag{5.6}$$

$$= \int_{[-L,L]^\Omega} \cdots \int \prod_p E_{p,\xi} \prod_{(i,j)\in\Omega} \varphi(\xi_{ij}) \mathrm{d}\xi_{ij} + O\Big(e^{-cL^{1/2}} (\log R)^{|\Omega|}\Big). \tag{5.7}$$

PROOF. Truncating the Fourier integral (5.2) at L, and using the decay estimate (5.3), we deduce that for every $(i, j) \in \Omega$, writing $z_{ij} = (1 + \xi_{ij})/\log R$,

$$\chi\Big(\frac{\log m_{ij}}{\log R}\Big) = \int_{-L}^{L} m_{ij}^{-z_{ij}} \varphi(\xi_{ij}) \mathrm{d}\xi_{ij} + O\Big(e^{-cL^{1/2}} m_{ij}^{-1/\log R}\Big).$$

Both terms in the right-hand side above are bounded by $O(m_{ij}^{-1/\log R})$, and therefore

$$\prod_{(i,j)\in\Omega} \chi\Big(\frac{\log m_{ij}}{\log R}\Big) = \int_{[-L,L]^\Omega} \cdots \int \prod_{(i,j)\in\Omega} m_{ij}^{-z_{ij}} \varphi(\xi_{ij}) \mathrm{d}\xi_{ij} + O\Big(e^{-cL^{1/2}} \prod_{(i,j)\in\Omega} m_{ij}^{-1/\log R}\Big).$$

Inserting this into (5.6), and exchanging sums and integrals, we obtain the expression

$$\int_{[-L,L]^\Omega} \cdots \int \sum_{(m_{ij})\in\mathbb{N}^\Omega}{}' \alpha(m_1, \dots, m_t) \prod_{(i,j)\in\Omega} \mu(m_{ij}) m_{ij}^{-z_{ij}} \prod_{(i,j)\in\Omega} \varphi(\xi_{ij}) \mathrm{d}\xi_{ij} \tag{5.8}$$

$$+ O\Big(e^{-cL^{1/2}} \sum_{(m_{ij})\in\mathbb{N}^\Omega}{}' \alpha(m_1, \dots, m_t) \prod_{(i,j)\in\Omega} m_{ij}^{-1/\log R}\Big).$$

By multiplicativity of $\alpha(m_1,\dots,m_t)$ in (m_{ij}), the main term in the above equals

$$\int_{[-L,L]^\Omega}\cdots\int\prod_p\sum_{(r_{ij})\in\{0,1\}^\Omega}(-1)^{\sum_{(i,j)\in\Omega}r_{ij}}\alpha(p^{r_1},\dots,p^{r_t})p^{-\sum_{(i,j)\in\Omega}r_{ij}z_{ij}}\prod_{(i,j)\in\Omega}\varphi(\xi_{ij})\mathrm{d}\xi_{ij},$$

where $r_i=\max(r_{i1},r_{i2})$. By (5.4) and reindexing by $B=\{(i,j):r_{ij}=1\}$, this equals

$$\int_{[-L,L]^\Omega}\cdots\int\prod_p E_{p,\xi}\prod_{(i,j)\in\Omega}\varphi(\xi_{ij})\mathrm{d}\xi_{ij}.$$

By similar considerations, the error term in (5.8) is

$$\ll e^{-cL^{1/2}}\prod_p\sum_{B\subset\Omega}\alpha(p,B)p^{-|B|/\log R}.$$

Since $\alpha(p,B)\leqslant p^{-1}$ for $B\neq\varnothing$ by Proposition 5.6, this error is further bounded by

$$e^{-cL^{1/2}}\prod_p\Big(1+\frac{|\Omega|}{p^{1+1/\log R}}\Big)\asymp e^{-cL^{1/2}}\prod_p\Big(1-\frac{1}{p^{1+1/\log R}}\Big)^{-|\Omega|}.$$

This last product equals $\zeta(1+\frac{1}{\log R})^{|\Omega|}$, and applying the elementary estimate $\zeta(s)=\frac{1}{s-1}+O(1)$ for $\mathrm{Re}(s)>0$, we see that the error is $\ll e^{-cL^{1/2}}(\log R)^{|\Omega|}$. □

From now on, we let $L\geqslant 1$ denote a truncation parameter, ξ denote an arbitrary real in $[-L,L]^\Omega$, and we keep the implicit notation $z_{ij}=(1+i\xi_{ij})/\log R$. From Proposition 5.6, we expect that, for large p, the main contribution to the sum defining $E_{p,\xi}$ in (5.5) comes from vertical sets B. It is then natural to approximate $E_{p,\xi}$ by the following Euler factor corresponding to a certain product (5.12) of zeta functions.

DEFINITION 5.9 (Auxiliary Euler factor). *We let*[7]

$$E'_{p,\xi}=\prod_{B\ \mathit{vertical}}\Big(1-p^{-1-\sum_B z_{ij}}\Big)^{-(-1)^{|B|}}. \tag{5.9}$$

The key estimates we need are the following.

[7] We write $\sum_B z_{ij}$ as short for $\sum_{(i,j)\in B}z_{ij}$.

PROPOSITION 5.10 (Euler factor estimates). *We have, uniformly in p,*

$$E_{p,\xi} = \begin{cases} 1 & \text{if } p \leqslant \omega, \\ (1 + O(p^{-2}))E'_{p,\xi} & \text{if } p > \omega. \end{cases}$$

Assuming further that $1 \leqslant L \leqslant \frac{c \log R}{\log \omega}$, we have, uniformly in $p \leqslant \omega$,

$$E'_{p,\xi} = \Big(1 + O\Big(\frac{L \log p}{p \log R}\Big)\Big) \cdot \Big(1 - \frac{1}{p}\Big)^t.$$

PROOF. We first observe that $|p^{-\sum_B z_{ij}}| = p^{-|B|/\log R} \leqslant 1$ for all p and $B \subset \Omega$. Now for $p \leqslant \omega$, we have $\alpha(p, B) = 0$ for all $B \neq \varnothing$ by Proposition 5.6, and therefore $E_{p,\xi} = 1$. For $p > \omega$, inserting the bounds of Proposition 5.6 into the definition (5.5) of $E_{p,\xi}$, we see that $E_{p,\xi}$ has an asymptotic expansion of the form

$$1 + \sum_{B \text{ vertical}} (-1)^{|B|} p^{-1-\sum_B z_{ij}} + O(p^{-2}), \tag{5.10}$$

which in particular is more than $1/2$ since ω is assumed to be large enough with respect to d, t. Using the same estimates in the product (5.9), we see that $E'_{p,\xi}$ also has an asymptotic expansion of the form (5.10), which yields the first estimate.

Since $1 \leqslant L \leqslant \frac{c \log R}{\log \omega}$, we have, for $p \leqslant \omega$, an approximation

$$p^{-\sum_B z_{ij}} = \exp\Big(O\Big(\frac{L \log p}{\log R}\Big)\Big) = 1 + O\Big(\frac{L \log p}{\log R}\Big).$$

Inserting this estimate in the product (5.9) defining $E'_{p,\xi}$, we obtain

$$E'_{p,\xi} = 1 + \Big(\sum_{B \text{ vertical}} (-1)^{|B|}\Big)\frac{1}{p} + O\Big(\frac{L \log p}{p \log R}\Big).$$

The second estimate then follows from computing

$$\sum_{B \text{ vertical}} (-1)^{|B|} = \sum_{i \in [t]} \Big(\sum_{B_i \subset [2]} (-1)^{|B_i|} - 1\Big) = -t. \tag{5.11}$$

□

Note that from the definition (5.9) of $E'_{p,\xi}$, we have

$$\prod_p E'_{p,\xi} = \prod_{B \text{ vertical}} \zeta\Big(1+\sum_B z_{ij}\Big)^{(-1)^{|B|}} \tag{5.12}$$

for every $\xi \in [-L,L]^\Omega$. It is then easy to estimate the size of this Euler product.

PROPOSITION 5.11 (Zeta function estimate). *Provided that* $1 \leqslant L \leqslant c\log R$, *we have*

$$\prod_p E'_{p,\xi} = \Big(1+O\Big(\frac{L}{\log R}\Big)\Big)\cdot(\log R)^{-t}\cdot \prod_{B \text{ vertical}} \Big(\sum_{(i,j)\in B}(1+i\xi_{ij})\Big)^{-(-1)^{|B|}}.$$

PROOF. From (5.12) and the estimate $\zeta(s) = \frac{1}{s-1}+O(1)$ for $\mathrm{Re}(s)>0$, we deduce that

$$\prod_p E'_{p,\xi} = \prod_{B \text{ vertical}} \Big(\frac{1}{\sum_B z_{ij}}+O(1)\Big)^{(-1)^{|B|}}.$$

From $|z_{ij}| \ll L/\log R$ we deduce that

$$\prod_p E'_{p,\xi} = \Big(1+O\Big(\frac{L}{\log R}\Big)\Big) \prod_{B \text{ vertical}} \Big(\sum_B z_{ij}\Big)^{-(-1)^{|B|}}.$$

The proposition follows from the definition $z_{ij} = (1+i\xi_{ij})/\log R$ and (5.11). □

We now have all the ingredients in hand to approximate the Euler product $\prod_p E_{p,\xi}$ efficiently.

PROPOSITION 5.12 (Euler product estimate). *Provided that* $1 \leqslant L \leqslant \frac{c\log R}{\log\omega}$, *we have*

$$\prod_p E_{p,\xi} = \Big(1+O\Big(\frac{1}{\omega\log\omega}+\frac{L\log\omega}{\log R}\Big)\Big)\cdot h_{R,W}^{-t}\cdot \prod_{B \text{ vertical}} \Big(\sum_{(i,j)\in B}(1+i\xi_{ij})\Big)^{-(-1)^{|B|}}.$$

PROOF. By Proposition 5.10 and Chebyshev's bounds, we have

$$\begin{aligned}\prod_p E_{p,\xi} &= \prod_{p>\omega}\Big(1+O\Big(\frac{1}{p^2}\Big)\Big)E'_{p,\xi}\\ &= \Big(1+O\Big(\frac{1}{\omega\log\omega}\Big)\Big)\prod_{p\leqslant\omega}{E'_{p,\xi}}^{-1}\prod_p E'_{p,\xi}.\end{aligned}\tag{5.13}$$

By the estimate of Proposition 5.10 on $E'_{p,\xi}$ and Chebyshev's bounds, we have

$$\prod_{p\leqslant\omega}{E'_{p,\xi}}^{-1} = \Big(1+O\Big(\frac{L\log\omega}{\log R}\Big)\Big)\Big(\frac{\phi(W)}{W}\Big)^{-t}.$$

Inserting finally the estimate of Proposition 5.11 into (5.13) concludes the proof. □

At this stage, the following sieve factors arise.

DEFINITION 5.13 (Sieve factor). *We let*

$$c_{\chi,2} = \iint_{\mathbb{R}^2}\frac{(1+i\xi)(1+i\xi')}{2+i(\xi+\xi')}\varphi(\xi)\varphi(\xi')\mathrm{d}\xi\mathrm{d}\xi'.$$

The last step is to replace the euler product $\prod_p E_{p,\xi}$ by $\prod_p E'_{p,\xi}$ in (5.7), and to extend the range of integration back to $\mathbb{R}$.

PROPOSITION 5.14 (Refolding integrals). *Provided that* $1\leqslant L\leqslant\frac{c\log R}{\log\omega}$, *we have*

$$\begin{aligned}&h_{R,W}^t\int_{[-L,L]^\Omega}\cdots\int\prod_p E_{p,\xi}\prod_{(i,j)\in\Omega}\varphi(\xi_{ij})\mathrm{d}\xi_{ij}\\ &= c_{\chi,2}^t + O\Big(e^{-cL^{1/2}}+\frac{1}{\omega\log\omega}+\frac{L\log\omega}{\log R}\Big).\end{aligned}\tag{5.14}$$

PROOF. By Proposition 5.12 and the Fourier decay (5.3), the expression (5.14) is equal to

$$\int_{\mathbb{R}^\Omega}\cdots\int\prod_{i\in[t]}\prod_{\substack{B_i\subset[2]\\B_i\neq\varnothing}}\Big(\sum_{j\in B_i}(1+i\xi_{ij})\Big)^{-(-1)^{|B_i|}}\prod_{j\in[2]}\varphi(\xi_{ij})\mathrm{d}\xi_{ij}+O\Big(\frac{1}{\omega\log\omega}+\frac{L\log\omega}{\log R}+e^{-cL^{1/2}}\Big).$$

To conclude observe that, by Fubini over $i\in[t]$, the main term above equals $c_{\chi,2}^t$. □

At this stage we quote [39, Lemma D.2], which provides an explicit formula for $c_{\chi,2}$.

LEMMA 5.15. *We have* $c_{\chi,2} = \int_0^\infty |\chi'(x)|^2 \mathrm{d}x$.

We may now combine the previous successive approximations to the original sum and optimize the parameter L to obtain Proposition 5.2.

Proof of Proposition 5.2. Let $P \geqslant 1$. Combining Propositions 5.5, 5.8 and 5.14, we see that the average $\mathbb{E}_{n\in[P]^d} \prod_{i\in[t]} \Lambda_{\chi,R,W}\big[\psi_i(n)\big]$ is equal to

$$c_{\chi,2}^t + O\bigg(e^{-cL^{1/2}}(\log R)^{O(1)} + \frac{1}{\omega \log \omega} + \frac{L \log \omega}{\log R} + \frac{R^{5t}}{P}\bigg),$$

provided that $L \leqslant \frac{c \log R}{\log \omega}$. Recall now that $\omega = c_0 \log N$. Assuming that $P \geqslant N^{c_1}$, we choose $L = C(\log\log N)^2$ and $R = N^{c_2/t}$ for a small $c_2 > 0$, so that

$$\mathbb{E}_{n\in[P]^d} \prod_{i\in[t]} \Lambda_{\chi,R,W}\big[\psi_i(n)\big] = c_{\chi,2}^t + O((\log N)^{-1+o(1)}). \tag{5.15}$$

By Lemma 5.15, we have $c_{\chi,2} > 0$ and therefore we may define a renormalized weight $\nu := c_{\chi,2}^{-1}\Lambda_{\chi,R,W}$, which satisfies the desired pseudorandomness asymptotic by (5.15), and which majorizes a constant multiple of $\lambda_{b,W}$ by Lemma 5.4. □

6. Quantitative pseudorandomness

The goal of this section is to transfer the previous pseudorandomness asymptotics over $\mathbb{Z}$ to the setting of a large cyclic group, and to show that pseudorandomness is preserved under certain averaging operations. We also state the generalized Von Neumann theorem of Green and Tao [39, Appendix C], in a quantified form. The relevant notion of pseudorandomness in our paper is the following.

DEFINITION 6.1 (Quantitative pseudorandomness). *Let $D, H \geqslant 1$ be parameters and let M be a prime. We say that $\nu : \mathbb{Z}_M \to \mathbb{R}^+$ is D-pseudorandom of level H when, for every affine system $\theta : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ of finite complexity such that*

$d, t, \|\dot{\theta}\| \leqslant D$,

$$\mathbb{E}_{n\in\mathbb{Z}_M^d}\nu\Big[\theta_1(n)\Big]\dots\nu\Big[\theta_t(n)\Big] = 1 + O_D\Big(\frac{1}{H}\Big).$$

We now let N denote an integer larger than some absolute constant, and as in the previous section we fix $\omega = c_0 \log N$ and $W = \prod_{p\leqslant\omega} p$. We also consider an embedding $[N] \hookrightarrow \mathbb{Z}_M$, where M is a prime larger than N. We are then interested in finding a pseudorandom majorant over $\mathbb{Z}_M$ for the function $\lambda_{b,W}$ from Definition 5.1, properly extended to a function on $\mathbb{Z}_M$. Precisely, given a function $f : \mathbb{Z} \to \mathbb{C}$ with support in $[N]$, we define an M-periodic function $\widetilde{f}$ at $n \in \mathbb{Z}$ by $\widetilde{f}(n) = f(n + \ell M)$, where ℓ is the unique integer such that $n + \ell M \in [M]$, and that function $\widetilde{f}$ may in turn be viewed as a function on $\mathbb{Z}_M$.

It is actually relatively simple to construct a pseudorandom majorant on $\mathbb{Z}_M$ from the one of Proposition 5.2, by cutting $\mathbb{Z}_M^d$ into small boxes as explained in [**36**, p. 527]. We rerun this argument here since we need to extract explicit error terms from it.

PROPOSITION 6.2 (Pseudorandom majorant over $\mathbb{Z}_M$). *Let $D \geqslant 1$. There exists a constant C_D such that if $N \geqslant C_D$ and $M \geqslant N$ is a prime, there exists a D-pseudorandom weight $\widetilde{\nu} : \mathbb{Z}_M \to \mathbb{R}^+$ of level $(\log N)^{1-o(1)}$ such that*

$$0 \leqslant \widetilde{\lambda}_{b,W} \ll_D \widetilde{\nu}.$$

PROOF. Consider an affine system $\theta : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ of finite complexity and such that $d, t, \|\dot{\theta}\| \leqslant D$. By Fact 4.12, we may consider θ as the reduction modulo M of an affine system $\psi : \mathbb{Z}^d \to \mathbb{Z}^t$ with norms $\|\psi\|_M = \|\theta\| \leqslant 2D$ and $\|\dot{\psi}\| = \|\dot{\theta}\| \leqslant D$. We let further implicit constants depend on D in the course of this proof.

Let ν be the weight from Proposition 5.2, and define $\widetilde{\nu} : \mathbb{Z}_M \to \mathbb{R}^+$ as above. Choosing another scale $P = M^{1/2}$, and duplicating the variable of averaging, we

obtain

$$\mathbb{E}_{n\in[M]^d}\prod_{i\in[t]}\widetilde{\nu}\Big[\psi_i(n)\Big]=\mathbb{E}_{m\in[M]^d}\mathbb{E}_{n\in[P]^d}\prod_{i\in[t]}\widetilde{\nu}\Big[\psi_i(m+n)\Big]+O(N^{-1/4}). \tag{6.1}$$

We call an integer m *good* when $\psi(m+[P]^d)\subset[M]^t+M\ell$ for some $\ell\in\mathbb{Z}^t$, and when that is not the case we say that m is *bad*. When m is good we have, with $\ell\in\mathbb{Z}^t$ as prescribed and by (5.1),

$$\begin{aligned}\mathbb{E}_{n\in[P]^d}\prod_{i\in[t]}\widetilde{\nu}\Big[\psi_i(m+n)\Big]&=\mathbb{E}_{n\in[P]^d}\prod_{i\in[t]}\nu\Big[\dot{\psi}_i(n)+(\psi_i(m)-M\ell_i)\Big]\\ &=1+O_D((\log N)^{-1+o(1)}).\end{aligned} \tag{6.2}$$

When m is bad, we have $\min_{i\in[t]}d(\psi_i(m),M\mathbb{Z})\leqslant\|\dot{\psi}\|P$ with respect to the canonical distance $d(x,y)=|x-y|$ on $\mathbb{R}$. Indeed, when that inequality does not hold, we have

$$\psi(m+]0,P[^d)\cap\{y\in\mathbb{R}^t:\exists i\in[t]\text{ such that }y_i\in M\mathbb{Z}\}=\varnothing,$$

and since $\psi(m+]0,P[^d)$ is connected it must be contained in one of the boxes $]0,M[^t+M\ell$, $\ell\in\mathbb{Z}^t$ (it is helpful to draw a picture at this point). We have thus proven that when m is bad, there exists $i\in[t]$ and $\ell_i\in\mathbb{Z}$ such that $\psi_i(m)\in\ell_iM+[-O(P),O(P)]$, and such an ℓ_i is necessarily $\ll 1+\|\psi\|_M\ll 1$. It is easy to check that the number of such $m\in[M]^d$ is $\ll PM^{d-1}=M^{d-1/2}$. Inserting the estimate (6.2) on good-boxes averages in (6.1), and neglecting the count of bad-boxes averages, we obtain the desired asymptotic. □

The notion of pseudorandomness is quite robust under averaging operations, as demonstrated by the following proposition, which is needed later on to majorize certain convolutions of $\lambda_{b,W}$.

PROPOSITION 6.3. *Let $D,H\geqslant 1$ be parameters and M be a prime. Suppose that $\nu:\mathbb{Z}_M\to\mathbb{R}^+$ is D-pseudorandom of level H, B is a symmetric subset of $\mathbb{Z}_M$*

and $\mu_B = (|B|/M)^{-1}1_B$. *Then* $\nu' = \frac{1}{2}(\nu + \nu * \mu_B)$ *is also* D*-pseudorandom of level* H.

PROOF. Consider an affine system $\theta : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ of finite complexity such that $d, t, \|\dot{\theta}\| \leqslant D$. Let $\nu^{(0)} = \nu$ and $\nu^{(1)} = \nu * \mu_B$, so that $\nu^{(\varepsilon)}(x) = \mathbb{E}_{y\in B}\nu(x + \varepsilon y)$ for every $\varepsilon \in \{0, 1\}$ and $x \in \mathbb{Z}_M$. Therefore

$$\begin{aligned} S &:= \mathbb{E}_{n\in\mathbb{Z}_M^d} \tfrac{\nu^{(0)}+\nu^{(1)}}{2}\big[\theta_1(n)\big] \cdots \tfrac{\nu^{(0)}+\nu^{(1)}}{2}\big[\theta_t(n)\big] \\ &= \mathbb{E}_{\varepsilon\in\{0,1\}^t}\mathbb{E}_{n\in\mathbb{Z}_M^d}\nu^{(\varepsilon_1)}\big[\theta_1(n)\big] \cdots \nu^{(\varepsilon_t)}\big[\theta_t(n)\big] \\ &= \mathbb{E}_{\varepsilon\in\{0,1\}^t}\mathbb{E}_{y\in B^t}\mathbb{E}_{n\in\mathbb{Z}_M^d}\nu\big[\theta_1(n) + \varepsilon_1 y_1\big] \cdots \nu\big[\theta_t(n) + \varepsilon_t y_t\big]. \end{aligned}$$

For every $\varepsilon \in \{0, 1\}^t$ and $y \in B^t$, the system $(\theta_i + \varepsilon_i y_i)_{1\leqslant i\leqslant t}$ has same linear part as $(\theta_i)_{1\leqslant i\leqslant t}$. Since ν is D-pseudorandom of level H, we have $S = 1 + O_D(H^{-1})$ as desired. □

We now quote the generalized Von Neumann theorem of Green and Tao [**39**, Appendix C]. It is simple to quantify the error term in that result in terms of the level of pseudorandomness of the weight.

THEOREM 6.4 (Generalized Von Neumann theorem). *Let* $d, t, Q, H \geqslant 1$ *and* $s \geqslant 0$ *be parameters, and let* $i \in [t]$ *be an indice. There exists a constant* D *depending on* d, t, Q *such that the following holds. Suppose that* $M > D$ *is a prime and* $\theta : \mathbb{Z}_M^d \to \mathbb{Z}_M^t$ *is an affine system of finite complexity in exact* s*-normal form at* i*, and such that* $\|\dot{\theta}\| \leqslant Q$. *Suppose also that* $\nu : \mathbb{Z}_M \to \mathbb{R}^+$ *is* D*-pseudorandom of level* H*, and* $f_1, \dots, f_t : \mathbb{Z}_M \to \mathbb{R}$ *are functions such that* $|f_j| \leqslant \nu$ *for every* $j \in [t]$. *Then we have*

$$\Big|\mathbb{E}_{n\in\mathbb{Z}_M^d} f_1\big[\theta_1(n)\big] \cdots f_t\big[\theta_t(n)\big]\Big|^{2^{s+1}} \leqslant \|f_i\|_{U^{s+1}(\mathbb{Z}_M)}^{2^{s+1}} + O_D(H^{-1}).$$

PROOF. Up to relabeling the f_j and θ_j, we may assume that $i = 1$. Up to permutating the base vectors, we may also assume that the set J_1 from Definition 4.3 is equal to $[s + 1]$. It then suffices to apply [**39**, Proposition 7.1"], whose proof

invokes twice the pseudorandomness condition of Definition 6.1, under the name "linear forms condition". Note that the argument there requires a change of variable $(x_1, \dots, x_{s+1}, y) \mapsto (c_1^{-1}x_1, \dots, c_{s+1}^{-1}x_{s+1}, y)$ with respect to the decomposition $\mathbb{Z}_M^d = \mathbb{Z}_M^{s+1} \times \mathbb{Z}_M^{d-(s+1)}$, where $c_k = \dot{\theta}_1(e_k)$. The condition $M > D \geqslant \|\dot{\theta}\|$ ensures that this is possible, however the new forms involved may have large size, potentially not bounded in terms of $\|\dot{\theta}\|$. Fortunately, it can be verified that making the change of variables $x_i \mapsto c_i c_{s+1} x_i$, $1 \leqslant i \leqslant s+1$ before each application of the linear forms condition in the proof of [**39**, Proposition 7.1"] converts the systems of forms under consideration back into sytems of bounded size. (Here we elaborated slightly on the footnote at the bottom of [**39**, p. 1822]). □

7. Translation-invariant equations in the primes

In this Section, we prove Theorem 1.2. Our two main tools are the transference principle of Helfgott and de Roton [**50**], including Naslund's [**65**] refinement thereof, and the relative generalized Von Neumann theorem of Green and Tao, in the quantitative form obtained in the previous section. These two tools together transfer the problem of finding a complexity-one pattern in the primes, to that of finding one in the integers, and to finish the proof we simply apply our extension of Shao's result derived in Section 8.

We now formally begin the proof of Theorem 1.2. We start with a standard preliminary reduction, the W-trick, which allows us to consider subsets of an arithmetic progression of modulus W in the primes instead.

THEOREM 7.1 (Theorem 1.2 in W-tricked primes). *Let $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ be a translation-invariant matrix of rank r and complexity one. There exists a constant C depending at most on r, t, V such that the following holds. Let $W = \prod_{p\leqslant\omega} p$, where $\omega = c_0 \log N$ with $c_0 \in [\frac{1}{4}, \frac{1}{2}]$, and let $b \in \mathbb{Z}$ such that $(b, W) = 1$. Suppose*

that A is a subset of $[N]$ such that $b + W \cdot A \subset \mathcal{P}$ and

$$|A| = \alpha(W/\phi(W))(\log N)^{-1}N,$$
$$\alpha \geqslant C(\log\log N)^{-1/25t}.$$

Then there exists $\mathbf{y} \in A^t$ with distinct coordinates such that $V\mathbf{y} = 0$.

Proof that Theorem 7.1 implies Theorem 1.2.

Consider a subset A of $\mathcal{P}_N$ of density α; we may certainly assume that $\alpha \geqslant CN^{-1/4}$, and in particular that N is large enough. Let $W = \prod_{p \leqslant \omega} p$, where $\omega = \frac{1}{4}\log N$, and let $N' = \lfloor N/W \rfloor = N^{3/4+o(1)}$ (by the prime number theorem) be another scale. By [**50**, Lemma 2.1], there exists $(b, W) = 1$ such that $A' = \{n \in [N'] : b + Wn \in A\}$ has size $\gg \alpha(W/\phi(W))(\log N')^{-1}N'$. Note that $\omega \sim \frac{1}{3}\log N'$ as $N \to \infty$, and since $b + W \cdot A' \subset A$, every solution $\mathbf{y} \in (A')^t$ to $V\mathbf{y} = 0$ with distinct coordinates induces one in A^t, by translation-invariance and homogeneity. Applying then Theorem 7.1 to $A' \subset [N']$ concludes the proof. □

From now on, we work under the hypotheses of Theorem 7.1. First, we consider an integer $N \geqslant 1$ and a constant $c_0 \in [\frac{1}{4}, \frac{1}{2}]$, and we fix

$$W = \prod_{p \leqslant \omega} p, \qquad \omega = c_0 \log N, \qquad b \in \mathbb{Z} : (b, W) = 1.$$

We then consider a subset $A \subset [N]$ such that $b + W \cdot A \subset \mathcal{P}$ and

$$|A| = \alpha \frac{W}{\phi(W)}(\log N)^{-1} \cdot N.$$

Accordingly, we define the normalized indicator function of A by

$$\lambda_A = \frac{\phi(W)}{W}(\log N) \cdot 1_A.$$

With this normalization, we have $\mathbb{E}\lambda_A = \alpha$ and, by comparison with Definition 5.1,

$$0 \leqslant \lambda_A \leqslant \lambda_{b,W}.$$

Secondly, we fix a translation-invariant matrix $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ of complexity one, and without loss of generality we may assume that $t \geqslant 3$ and V has no zero columns in proving Theorem 7.1. Via Propositions 4.8 and 4.9, we can choose a linear parametrization $\psi : \mathbb{Z}^d \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}_{\mathbb{Q}}(V)$ in exact 1-normal form over $\mathbb{Z}$ at every $i \in [t]$. We assume from now on that N is large enough with respect to d, t, ψ, V, and we let further implicit and explicit constants depend on those parameters. We will need to consider functions with support in $[-2N, 2N]_{\mathbb{Z}}$, and to analyze those we embed $[-2N, 2N]_{\mathbb{Z}}$ in a large cyclic group $\mathbb{Z}_M$, where M is a prime between $4(\|V\| + 1) \cdot N$ and $8(\|V\| + 1) \cdot N$ chosen via Bertrand's postulate. By Fact 4.11, the linear map ψ reduces modulo M to a linear map $\theta : \mathbb{Z}_M^d \twoheadrightarrow \mathrm{Ker}_{\mathbb{Z}_M}(V)$ in exact 1-normal form over $\mathbb{Z}_M$ at every $i \in [t]$, and such that $\|\theta\| = \|\psi\|$; we work exclusively with that map from now on.

Given a function $f : \mathbb{Z} \to \mathbb{C}$ with support in $[-2N, 2N]$, we define an M-periodic function $\breve{f}(n) = 0$ at $n \in \mathbb{Z}$ by $\breve{f}(n) = f(n + \ell M)$, where ℓ is the unique integer such that $n + \ell M \in [-M/2, M/2]_{\mathbb{Z}}$, and $\breve{f}$ may then be considered as a function on $\mathbb{Z}_M$. When f has support in $[N]$, as is the case for $\lambda_{b,W}$, this coincides with the definition of $\widetilde{f}$ from Section 6. To alleviate the notation, we now identify functions $f : \mathbb{Z} \to \mathbb{C}$ with support in $[-2N, 2N]$ with their periodic counterpart $\breve{f}$. Most of the analysis we do next takes place on $\mathbb{Z}_M$, and Fourier transforms, convolutions, L^p and U^k norms are normalized accordingly. With these notations in place, we now work with the following pattern-counting operator.

DEFINITION 7.2. *We define the operator T on functions $f_1, \dots, f_t : \mathbb{Z}_M \to \mathbb{R}$ by*

$$T(f_1, \dots, f_t) = \mathbb{E}_{n\in\mathbb{Z}_M^d} f_1\big[\theta_1(n)\big] \dots f_t\big[\theta_t(n)\big].$$

If need be, we can always return to averages over $\mathbb{Z}$ via the following observation.

LEMMA 7.3. *For functions $f_1, \dots, f_t : \mathbb{Z}_M \to \mathbb{R}$ with support in $[-2N, 2N]$, we have*

$$T(f_1, \dots, f_t) = M^{-(t-r)} \sum_{\substack{y \in [-2N,2N]^t_{\mathbb{Z}} : \\ Vy = 0}} f_1(y_1) \dots f_t(y_t).$$

PROOF. Since θ is a surjection onto $\mathrm{Ker}_{\mathbb{Z}_M}(V)$, and the fibers $\#\{x \in \mathbb{Z}_M^d : \theta(x) = y\}$ have uniform size when y ranges over $\mathrm{Ker}_{\mathbb{Z}_M}(V)$, we have

$$\begin{aligned} T(f_1, \dots, f_t) &= \mathbb{E}_{y \in \mathbb{Z}_M^t : Vy = 0} f_1(y_1) \dots f_t(y_t) \\ &= M^{-(t-r)} \textstyle\sum_{y \in \mathbb{Z}_M^t : Vy = 0} f_1(y_1) \dots f_t(y_t). \end{aligned}$$

Since the f_i have support in $[-2N, 2N]$, we may restrict the summation to $y \in [-2N, 2N]^t_{\mathbb{Z}}$, and since $M > 2\|V\|N$, the identity $Vy = 0$ holds in $\mathbb{Z}$ for such y. □

We now introduce two parameters $\delta \in (0, 1]$ and $\varepsilon \in (0, c]$. We also fix an auxiliary Bohr set of $\mathbb{Z}_M$ (see Definition 8.3) defined by

$$\begin{aligned} \Gamma &= \{r \in \mathbb{Z}_M : |\widehat{\lambda}_A(r)| \geqslant \delta\} \cup \{1\}, \\ B &= B(\Gamma, \varepsilon). \end{aligned}$$

The presence of 1 in the frequency set guarantees that the Bohr set is contained in an interval $[-\varepsilon M, \varepsilon M]$. As is common in the transference literature for three-term arithmetic progressions [**30**, **34**, **50**, **65**], we work with a smooth approximation of λ_A, namely the convolution over $\mathbb{Z}$ given by

$$\lambda'_A = \lambda_A * \lambda_B,$$

where $\lambda_B = |B|^{-1} 1_B$. Provided that ε is small enough, we see that the support of λ'_A is contained in $[-2N, 2N]$. Since $M > 2N$, we may also consider $\lambda'_A : \mathbb{Z}_M \to \mathbb{R}$

as the normalized convolution over $\mathbb{Z}_M$ given by

$$\lambda'_A = \lambda_A * \mu_B, \tag{7.1}$$

where $\mu_B = (|B|/M)^{-1}1_B$. To show that λ'_A is close to λ_A in a Fourier ℓ^4 sense, we need to call on the restriction estimates of Green and Tao [34], themselves based on an envelopping sieve of Ramaré and Ruzsa [68]; these estimates were in turn adapted to the case of a large modulus ω by Helfgott and de Roton [50].

PROPOSITION 7.4. *We have* $\|\lambda_A - \lambda'_A\|_{U^2} \ll \varepsilon^{1/4} + \delta^{1/4}$.

PROOF. By [50, Lemma 2.2], we have $\sum_r |\widehat{\lambda}_A(r)|^q \ll_q 1$ for any $q > 2$. Therefore,

$$\begin{aligned}\|\lambda_A - \lambda'_A\|_{U^2}^4 &= \sum_r |\widehat{\lambda}_A(r)|^4 |1 - \widehat{\mu}_B(r)|^4 \\ &\ll \varepsilon \sum_{r:\, |\widehat{\lambda}_A(r)| \geqslant \delta} |\widehat{\lambda}_A(r)|^4 + \delta \sum_{r:\, |\widehat{\lambda}_A(r)| \leqslant \delta} |\widehat{\lambda}_A(r)|^3 \\ &\ll \varepsilon + \delta,\end{aligned}$$

where we used the fact that $|1-\widehat{\mu}_B(r)| = |\mathbb{E}_{x\in B}(1-e_N(rx))| \leqslant 2\pi\varepsilon$ for all $r \in \Gamma$. □

The structure of our argument is now as follows: we compare the counts $T(\lambda_A, \dots, \lambda_A)$ and $T(\lambda'_A, \dots, \lambda'_A)$, which we expect to be close by Proposition 7.4 and the heuristic that "U^2 norm controls complexity one averages".

REMARK 7.5 (Multilinear expansion). *By multlinearity,*

$$T(\lambda_A, \dots, \lambda_A) = T(\lambda'_A, \dots, \lambda'_A) + \sum T(*, \dots, \lambda_A - \lambda'_A, \dots, *). \tag{7.2}$$

where the sum is over $2^t - 1$ *terms and the stars stand for functions equal to* λ'_A *or* $\lambda_A - \lambda'_A$.

To estimate the main term in (7.2), that is, $T(\lambda'_A, \dots, \lambda'_A)$, we invoke a key transference estimate of Helfgott and de Roton [50], which essentially allows us to consider λ'_A as a subset of the integers of density α^2. It is further possible,

by a result of Naslund[8] [65], to obtain an exponent $1 + o(1)$ instead of 2, and we choose to work with that more efficient version, even though it is possible to derive Theorem 1.2 with a smaller exponent without it. This is because we wish to exhibit that our argument preserves the exponent in Szemerédi-type theorems in the integers, in the sense of Proposition 7.9 below.

PROPOSITION 7.6. *Suppose that* $\delta^{-4}\log\varepsilon^{-1} \leqslant c\log N$. *Then for any* $\kappa > 0$, *the level set* $A' = \{\lambda'_A \geqslant \alpha/2\}$ *has density* $\gg_\kappa \alpha^{1+\kappa}$ *in* $\mathbb{Z}_M$.

PROOF. Recalling (7.1), we see that $\mathbb{E}\lambda'_A = \mathbb{E}\lambda_A = \alpha$. By Selberg's sieve or the restriction estimate used in the proof of Proposition 7.4, we have

$$\#\{r : |\widehat{\lambda}_A(r)| \geqslant \delta\} \leqslant \delta^{-4}\|\widehat{\lambda}_A\|_4^4 \ll \delta^{-4},$$

and therefore $|B| \geqslant \varepsilon^{|\Gamma|}N \geqslant N^{1/2}$ under our assumptions on ε and δ. By [65, Proposition 2], we deduce that $\|\lambda'_A\|_p \ll_p 1$ for any even $p \geqslant 4$, and the proposition then follows from a simple bootstrapping argument [65, Lemma 6]. □

Applying our statistical, complexity-one extension of Shao's result in the integers, we can now obtain a lower bound on the average of λ'_A over ψ-configurations.

PROPOSITION 7.7 (Main term). *Suppose that* $\delta^{-4}\log\varepsilon^{-1} \leqslant c\log N$. *We have*

$$T(\lambda'_A, \dots, \lambda'_A) \geqslant \exp\left[-C_\kappa \alpha^{-24t-\kappa}\right]$$

for every $\kappa > 0$.

PROOF. Consider the level set $A' = \{\lambda'_A \geqslant \alpha/2\}$ contained in the support of λ'_A, and therefore in $[-2N, 2N]$. Since $\lambda'_A \geqslant (\alpha/2)\cdot 1_{A'}$, we have

$$T(\lambda'_A, \dots, \lambda'_A) \geqslant (\alpha/2)^t T(1_{A'}, \dots, 1_{A'}).$$

[8] Here we implicitely refer to the first version of Naslund's preprint, because the argument there is simpler, and we do not seek very sharp bounds on the exponent.

By Proposition 7.6, we know that A' has density $\gg_\kappa \alpha^{1+\kappa}$ in $[-2N, 2N]$ for any $\kappa > 0$. Invoking Lemma 7.3, and applying Proposition 8.1 to $A' \subset [-2N, 2N]$, we obtain

$$T(1_{A'}, \dots, 1_{A'}) = M^{-(t-r)} \#\{y \in (A')^t : Vy = 0\} \geqslant \exp\Big[- C_\kappa \alpha^{-(1+\kappa)24t} \Big].$$

□

On the other hand, the averages from (7.2) involving a difference $\lambda_A - \lambda'_A$ are bounded via the generalized Von Neumann theorem of Section 6.

PROPOSITION 7.8 (Error terms). *Suppose that $f_1, \dots, f_t$ are functions all equal to λ'_A or $\lambda_A - \lambda'_A$, with at least one of them equal to $\lambda_A - \lambda'_A$. Then*

$$|T(f_1, \dots, f_t)| \ll \varepsilon^{1/4} + \delta^{1/4} + (\log N)^{-\frac{1}{4}+o(1)}.$$

PROOF. We consider $i \in [t]$ such that $f_i = \lambda_A - \lambda'_A$. Let $Q = \|\dot\theta\|$ and let $D = D_{d,t,Q}$ be the constant from Proposition 6.4. By Proposition 6.2, and since we assumed N to be large enough with respect to d, t, θ, there exists a D-pseudorandom weight $\nu : \mathbb{Z}_M \to \mathbb{R}^+$ of level $(\log N)^{1-o(1)}$ such that

$$0 \leqslant \lambda_A \leqslant \lambda_{b,W} \ll \nu.$$

Let $\nu' = \frac{1}{2}(\nu + \nu * \mu_B)$, so that $|\lambda'_A| \ll \nu'$ and $|\lambda_A - \lambda'_A| \ll \nu'$. By Proposition 6.3, ν' is also D-pseudorandom of level $(\log N)^{1-o(1)}$.

Recall now that ψ is in exact 1-normal form at i. Applying Proposition 6.4 with $s = 1$ to the functions $f_1, \dots, f_t$ (divided by a certain large constant), and inserting the estimates of Proposition 7.4, we obtain the desired bound. □

At this point we need only collect together the bounds on the main term and the error terms in (7.2) to finish the proof of Theorem 1.2, which we have previously reduced to proving Theorem 7.1.

Proof of Theorem 7.1. Starting from the multilinear expansion (7.2), and inserting the bounds from Propositions 7.7 and 7.8, we obtain

$$T(\lambda_A, \dots, \lambda_A) \geqslant \exp[-C_\kappa \alpha^{-24t-\kappa}] - O\Big(\varepsilon^{1/4} + \delta^{1/4} + (\log N)^{-\frac{1}{4}+o(1)}\Big),$$

whenever, say, $\varepsilon^{-1}, \delta^{-1} \leqslant c(\log N)^{1/8}$. Choose now $\varepsilon = \delta = \exp[-C'_\kappa \alpha^{-24t-\kappa}]$ (for a large C'_κ), and assume that $\alpha \geqslant C_\kappa (\log\log N)^{-1/(24t+\kappa)}$. This ensures that the conditions on ε and δ are satisfied, and that we have a lower bound

$$T(\lambda_A, \dots, \lambda_A) \geqslant \exp[-C'_\kappa \alpha^{-24t-\kappa}].$$

By Lemma 7.3 and since $\lambda_A \leqslant (\log N) 1_A$, we then have

$$\#\{y \in A^t : Vy = 0\} \geqslant \exp\Big[-C_\kappa \alpha^{-24t-\kappa}\Big] \cdot N^{t-r} (\log N)^{-t}.$$

On the other hand, by Lemma 4.10, the number of $y \in [N]^t$ with two identical coordinates and such that $Vy = 0$ is $\ll N^{t-r-1}$. Choosing now $\kappa = t$ for aesthetic reasons, and given the range of density under consideration, we are therefore ensured to find at least one non-trivial solution. □

As claimed before, our argument allows for a slightly more general statement than Theorem 1.2. Indeed, the following can be obtained by a suitable Varnavides argument and by inserting the resulting analog of Proposition 8.1 in our proof.

THEOREM 7.9. *Suppose that $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ is a translation-invariant matrix of rank r and complexity one, and let $\gamma > 0$ be a parameter. Assume that $V\mathbf{y} = 0$ has a distinct-coordinates solution $\mathbf{y} \in A^t$ for every subset A of $[N]$ of density at least*

$$C(\log N)^{-\gamma}.$$

Then such a solution also exists for every subset A of $\mathcal{P}_N$ of density at least

$$C_\varepsilon (\log\log N)^{-\gamma+\varepsilon},$$

for any $\varepsilon > 0$.

This being said, we have not tried to optimize the exponent $1/24t$ in Corollary 8.2, or the exponent in Theorem 1.2 that follows from it. This is because this exponent is likely not optimal, and far from comparable in quality with Sanders' [**81**] bounds for Roth's theorem, due to the repeated applications of Cauchy-Schwarz in Section 8.

8. Appendix: Translation-invariant equations in the integers

The purpose of this section is to derive an extension of a result of Shao [**91**] to arbitrary systems of complexity one, and with a count of the multiplicity of pattern occurences. The structure of our proof is similar to Shao's, and it relies in particular in the key local inverse U^2 theorem proved there (Proposition 8.12 below). However, certain added technicalities arise when handling arbitrary systems: the most significant of those is addressed by Proposition 8.11 below.

PROPOSITION 8.1. *Let* $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ *be a translation-invariant matrix of rank* r *and complexity one. Suppose that* A *is a subset of* $[-N, N]_{\mathbb{Z}}$ *of density* α. *Then*

$$\#\{\mathbf{y} \in A^t : V\mathbf{y} = 0\} \geqslant \exp\left[-C\alpha^{-24t}\right] \cdot N^{t-r},$$

for a constant $C > 0$ *depending at most on* r, t, V.

Although we only need the result above for the transference argument of Section 7, we record the following consequence, since it may be of independent interest.

COROLLARY 8.2. *Let* $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ *be a translation-invariant matrix of rank* r *and complexity one. There exists a constant* $C > 0$ *depending at most on* r, t, V *such that, if* A *is a subset of* $[N]$ *of density at least* $C(\log N)^{-1/24t}$, *there exists a solution* $\mathbf{y} \in A^t$ *to* $V\mathbf{y} = 0$ *with distinct coordinates.*

PROOF. By Lemma 4.10, the number of $\mathbf{y} \in [N]^t$ with two equal coordinates such that $V\mathbf{y} = 0$ is at most $O(N^{t-r-1})$. The result then follows from Proposition 8.1, since we assumed that $\alpha \geqslant C(\log N)^{-1/24t}$. □

We now fix a translation-invariant matrix $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ of rank r, and for the purpose of proving Proposition 8.1, we may assume without loss of generality that $t \geqslant 3$ and V has no zero columns. By Propositions 4.8 and 4.9, we may choose a linear parametrization $\varphi : \mathbb{Z}^{q+1} \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}_{\mathbb{Q}}(V)$ of the form $\varphi(x_0, x) = x_0\mathbf{1} + \psi(x)$, where $\psi : \mathbb{Z}^q \to \mathbb{Z}^t$ is in exact 1-normal form at every $i \in [t]$. We have traded the letter d for q here because the former is too precious as the dimension of a Bohr set. Writing $\psi_i(x) = a_{i1}x_1 + \cdots + a_{iq}x_q$, we define the sets of non-zero coefficients $\Xi_i = \{a_{ij} \neq 0,\ j \in [q]\}$ and $\Xi = \cup_{i\in[t]}\Xi_i$, so that we have $|a| \leqslant \|\varphi\|$ for every $a \in \Xi$.

We also consider a fixed integer N from the statement of Proposition 8.1, which should be thought of as quite large. As usual, we choose to carry out our Fourier analysis over a cyclic group $\mathbb{Z}_M$ on a slightly larger scale; to be precise, via Bertrand's postulate we pick a prime M such that $\|\varphi\| \cdot 2N < M \leqslant \|\varphi\| \cdot 4N$. Finally, throughout this section the letters c and C denote positive constants which are chosen, respectively, small or large enough with respect to q, t and φ. While we do not attempt to track the dependency of our parameters on $\|\varphi\|$, we sometimes use this quantity to illustrate our argument.

We now recall the basics of Bohr sets and regularity calculus, which can be found in many places [27, 35, 52]. We speed up this process as this material is utterly standard and our notation is consistent with the litterature.

DEFINITION 8.3. *A Bohr set of frequency set $\Gamma \subset \mathbb{Z}_M$ and radius $\delta > 0$ is*

$$B(\Gamma, \delta) = \{x \in \mathbb{Z}_M : \|\tfrac{xr}{M}\| \leqslant \delta \quad \forall r \in \Gamma\},$$

and its dimension d is defined by $d = |\Gamma|$. We often let the parameters Γ, δ, d be implicitely defined whenever we introduce a Bohr set B. The ρ-dilate $B_{|\rho}$ of a Bohr set B is defined by $B(\Gamma, \delta)_{|\rho} = B(\Gamma, \rho\delta)$, and given two Bohr sets B, B' we write $B' \leqslant_\rho B$ when $B' \subset B_{|\rho}$. Finally, we say that B is regular when, for every

$0 < \rho \leqslant 2^{-6}/d$,

$$(1 - 2^6\rho d)|B| \leqslant |B_{|1\pm\rho}| \leqslant (1 + 2^6\rho d)|B|.$$

We also recall standard size estimates on Bohr sets, as well as Bourgain's regularization lemma. In our later argument, all Bohr sets will be picked regular.

FACT 8.4. *Suppose that B is a Bohr set of dimension d and radius δ, and $\rho \in (0, 1]$. Then*

$$|B| \geqslant \delta^d M \quad \textit{and} \quad |B_{|\rho}| \geqslant (\rho/2)^{2d}|B|.$$

Given any Bohr set B, there exists $c \in [\frac{1}{2}, 1]$ such that $B_{|c}$ is regular.

In practice, regularity is used in the following form, close in spirit to [35, Lemma 4.2]. When we argue "by regularity" in a proof, we implicitely invoke these estimates.

FACT 8.5 (Regularity calculus). *Let $f : \mathbb{Z}_M \to [-1, 1]$ and suppose that B is a regular d-dimensional Bohr set, $X' \subset B_{|\rho}$ is another set and $x' \in B_{|\rho}$, where $\rho \in (0, c/d\,]$. Then*

$$\mathbb{E}_{x\in x'+B} f(x) = \mathbb{E}_{x\in B} f(x) + O(\rho d),$$

$$\mathbb{E}_{x\in B} f(x) = \mathbb{E}_{x\in B, x'\in X'} f(x + x') + O(\rho d),$$

$$\mathbb{E}_{x\in B} 1(x \in B_{|1-\rho}) f(x) = \mathbb{E}_{x\in B} f(x) + O(\rho d).$$

Before proceeding further, we recall certain facts about Gowers box norms [39, Appendix B], which are present in disguise in Shao's argument [91]. For our argument, we only require the positivity of such norms, and two Cauchy-Schwarz-based inequalities. Strictly speaking, we could do without those norms, however they are useful to write averages over cubes in a more compact (if less intuitive)

form, and to expedite repeated applications of Cauchy-Schwarz. In the following definitions, we let X_1, X_2 denote arbitrary subsets of $\mathbb{Z}_M$.

DEFINITION 8.6 (Box scalar product and norm). *The box scalar product of a family of functions* $(h_\omega : X_1 \times X_2 \to \mathbb{R})_{\omega\in\{0,1\}^2}$ *is*

$$\langle (h_\omega)\rangle_{\square(X_1\times X_2)} = \mathbb{E}_{x^{(0)},x^{(1)}\in X_1\times X_2} \prod_{\omega\in\{0,1\}^2} h_\omega(x_1^{(\omega_1)}, x_2^{(\omega_2)}).$$

The box norm of a function $h : X_1 \times X_2 \to \mathbb{R}$ *is defined by* $\|h\|^4_{\square(X_1\times X_2)} = \langle (h)\rangle_{\square(X_1\times X_2)}$.

The first inequality we require is a box Van der Corput inequality implicit in [23, p. 161], while the second is the Gowers-Cauchy-Schwarz inequality [39, Lemma B.2].

FACT 8.7. *For* $h : X_1 \times X_2 \to \mathbb{R}$ *and* $(b_k : X_k \to [-1, 1])_{k\in\{1,2\}}$, *we have*

$$\left|\mathbb{E}_{x_1\in X_1, x_2\in X_2} h(x_1, x_2)b_1(x_1)b_2(x_2)\right| \leqslant \|h\|_{\square(X_1\times X_2)}. \tag{8.1}$$

For $(h_\omega : X_1 \times X_2 \to \mathbb{R})_{\omega\in\{0,1\}^2}$, *we have*

$$\left|\langle (h_\omega)\rangle_{\square(X_1\times X_2)}\right| \leqslant \prod_{\omega\in\{0,1\}^2} \|h_\omega\|_{\square(X_1\times X_2)}. \tag{8.2}$$

In our situation, we need a slight variant of the local U^2 norm defined in [91].

DEFINITION 8.8 (Twisted U^2 norm). *Let* $a, b \in \mathbb{Z}$ *and* $g : \mathbb{Z}_M \to \mathbb{R}$. *The* (a, b)*-twisted* U^2 *norm of* g *with respect to* X_1, X_2 *is*

$$\|g\|^4_{\boxtimes_{a,b}(X_1\times X_2)} = \mathbb{E}_{x^{(0)},x^{(1)}\in X_1\times X_2} \prod_{\omega\in\{0,1\}^2} g(ax_1^{(\omega_1)} + bx_2^{(\omega_2)}).$$

When $a = b = 1$ *we simply write* $\|g\|_{\boxtimes(X_1\times X_2)}$.

With these notations, the local Gowers norm of a function f with respect to sets X_0, X_1, X_2 as defined by Shao [91, Definition 3.1] is

$$\|f\|^4_{U^2(X_0,X_1,X_2)} = \mathbb{E}_{x_0\in X_0}\|f(x_0 + \cdot)\|^4_{\boxtimes(X_1\times X_2)}.$$

From now on we keep the suggestive "local Gowers norm" terminology, but we use the expression in the right-hand side for computational purposes.

We are now ready to start with the proof of Proposition 8.1. We introduce, for a system of Bohr sets $\mathbf{B} = (B_0, \dots, B_q)$, the multilinear operator on functions

$$T_{\mathbf{B}}(f_1, \dots, f_t) = \mathbb{E}_{x_0 \in B_0, \dots, x_q \in B_q} f_1\Big[\varphi_1(x)\Big] \dots f_t\Big[\varphi_t(x)\Big].$$

The next proposition then constitutes the first step of our density increment strategy, in which we deduce that a set A either possesses many φ-configurations, or it induces a large $T_{\mathbf{B}}$-average involving the balanced function of A. Here and in the following, we occasionally make superfluous assumptions on the Bohr sets involved, in order to facilitate the combination of intermediate propositions.

PROPOSITION 8.9 (Multilinear expansion). *Suppose that A is a subset of density α of a regular d-dimensional Bohr set $B = B_0$, and write $f_A = 1_A - \alpha 1_B$. Suppose also that $B_1, \dots, B_q$ are regular Bohr sets with $B_i \leqslant_\rho B_{i-1}$ for all $i \in [q]$, where $\rho \leqslant c/d$. Then either*

(i) *(Many patterns) $T_{\mathbf{B}}(1_A, \dots, 1_A) \geqslant \alpha^t/4$,*

(ii) *(Large T-average) or there exist functions $f_1, \dots, f_t : \mathbb{Z}_M \to [-1, 1]$ and $i \in [t]$ such that $f_i = f_A$ and $|T_{\mathbf{B}}(f_1, \dots, f_t)| \gg \alpha^t$.*

PROOF. First observe that, expanding $1_A = \alpha 1_B + f_A$ by multilinearity,

$$T_{\mathbf{B}}(1_A, \dots, 1_A) = T_{\mathbf{B}}(\alpha 1_B, \dots, \alpha 1_B) + \sum T_{\mathbf{B}}(*, \dots, f_A, \dots, *) \tag{8.3}$$

where the sum is over $2^t - 1$ terms and the stars stand for functions equal to $\alpha 1_B$ or f_A. By definition,

$$T_{\mathbf{B}}(\alpha 1_B, \dots, \alpha 1_B) = \alpha^t \mathbb{E}_{x_0 \in B} \mathbb{E}_{x \in B_1 \times \dots \times B_q} 1_B\Big[x_0 + \psi_1(x)\Big] \dots 1_B\Big[x_0 + \psi_t(x)\Big].$$

Restricting x_0 to lie in $B_{|1-\rho}$ with $\rho \leqslant c/\|\varphi\|d$, we are ensured that $x_0 + \psi_j(x) \in B$ for every $j \in [t]$ and $x \in B_1 \times \cdots \times B_q \subset B^q_{|\rho}$. By regularity, we thus have

$$\begin{aligned} T_{\mathbf{B}}(\alpha 1_B, \dots, \alpha 1_B) &= \alpha^t \Big(\mathbb{E}_{x_0 \in B} 1_{B_{|1-\rho}}(x_0) + O(\rho d) \Big) \\ &= (1 + O(\rho d))\alpha^t \\ &\geqslant \alpha^t/2. \end{aligned}$$

By (8.3), if we are not in the first case of the proposition, then by the pigeonhole principle there must exist a large average

$$\alpha^t \ll |T_{\mathbf{B}}(f_1, f_2, \dots, f_t)|$$

where one of the functions $f_i : \mathbb{Z}_M \to [-1, 1]$ is equal to f_A. □

The next step is to use the fact that (twisted) local Gowers norms control the count of φ-configurations, up to a small error. This is the analog for general systems of complexity 1 of Shao's [**91**, Proposition 4.1]; it is also very similar to Green and Tao's generalized Von Neumann theorem for bounded functions [**23**, Theorem 2.3].

PROPOSITION 8.10 (Large average implies large Gowers norm). *Let $\eta \in (0, 1]$ be a parameter, and suppose that $B_0, \dots, B_q$ are regular d-dimensional Bohr sets such that $B_i \leqslant_\rho B_{i-1}$ for all $i \in [q]$, where $\rho \leqslant c\eta^4/d$. Suppose that $f_1, \dots, f_t : \mathbb{Z}_M \to [-1, 1]$ are such that*

$$|T_{\mathbf{B}}(f_1, \dots, f_t)| \geqslant \eta.$$

Then for every $i \in [t]$, there exist $1 \leqslant k < \ell \leqslant q$ and $a, b \in \Xi_i$ such that

$$\mathbb{E}_{u_0 \in B_0} \|f_i(u_0 + \cdot)\|^4_{\boxtimes_{a,b}(B_k \times B_\ell)} \geqslant \eta/2.$$

PROOF. Let $i \in [t]$, and recall that ψ is in exact 1-normal form at i. We may therefore find indices $1 \leqslant k < \ell \leqslant q$ and a partition $[t] \setminus \{i\} = X_k \sqcup X_\ell$ into non-empty sets such that ψ_i depends on the variables x_k and x_ℓ, while for $j \in X_k$

(respectively $j \in X_\ell$), ψ_j depends at most on the variable x_k (respectively x_ℓ) among those two variables. We decompose vectors $x \in \mathbb{Z}^{q+1}$ accordingly as $x = (x_0, x_k, x_\ell, y)$ with $y \in \prod_{j \notin \{0,k,l\}} B_j$, and we may write $\psi_i(x_k, x_\ell, y) = a_k x_k + a_\ell x_\ell + \psi_i(0,0,y)$ with $a_k, a_\ell \in \Xi_i$. Then[9]

$$\eta \leqslant \Big| \mathbb{E}_{x_0 \in B_0, y \in (B_j)_{j \notin \{0,k,\ell\}}} \mathbb{E}_{x_k \in B_k, x_\ell \in B_\ell} f_i\Big[x_0 + \psi_i(x_k, x_\ell, y)\Big] \\ \times \prod_{j \in X_k} f_j\Big[x_0 + \psi_j(x_k, y)\Big] \prod_{j \in X_\ell} f_j\Big[x_0 + \psi_j(x_\ell, y)\Big]\Big|.$$

We may rewrite the averaged function as $h(x_k, x_\ell) b_k(x_k) b_\ell(x_\ell)$, where h, b_k, b_ℓ are functions depending on x_0, y and b_k, b_ℓ are bounded by 1. By Hölder's inequality, followed by the box Van der Corput inequality (8.1), we thus have

$$\begin{aligned} \eta^4 &\leqslant \Big(\mathbb{E}_{x_0 \in B_0, y \in (B_j)_{j \notin \{0,k,\ell\}}} \Big| \mathbb{E}_{x_k \in B_k, x_\ell \in B_\ell} h(x_k, x_\ell) b_k(x_k) b_\ell(x_\ell) \Big| \Big)^4 \\ &\leqslant \mathbb{E}_{x_0 \in B^{(0)}, y \in (B_j)_{j \notin \{0,k,\ell\}}} \Big| \mathbb{E}_{x_k \in B_k, x_\ell \in B_\ell} h(x_k, x_\ell) b_k(x_k) b_\ell(x_\ell) \Big|^4 \\ &\leqslant \mathbb{E}_{x_0 \in B_0, y \in (B_j)_{j \notin \{0,k,\ell\}}} \|h\|^4_{\square(B_k \times B_\ell)}. \end{aligned}$$

Unfolding the definition of the box norm, and by regularity on the variable x_0, we have

$$\begin{aligned} \eta^4 &\leqslant \mathbb{E}_{x_0 \in B_0, y \in (B_j)_{j \notin \{0,k,\ell\}}} \mathbb{E}_{x^{(0)}, x^{(1)} \in B_k \times B_\ell} \\ &\qquad \prod_{\omega \in \{0,1\}^2} f_i(x_0 + a_k x_k^{(\omega_k)} + a_\ell x_\ell^{(\omega_\ell)} + \psi_i(0,0,y)) \\ &= \mathbb{E}_{x_0 \in B_0} \mathbb{E}_{x^{(0)}, x^{(1)} \in B_k \times B_\ell} \prod_{\omega \in \{0,1\}^2} f_i(x_0 + a_k x_k^{(\omega_k)} + a_\ell x_\ell^{(\omega_\ell)}) + O(\rho d). \end{aligned}$$

Refolding the definition of the (a_k, a_ℓ)-twisted U^2 norm, this concludes the proof, provided that $\rho \leqslant c\eta^4/d$. □

We now wish to reduce the conclusion of the previous proposition to the situation where $a = b = 1$, that is, when f_A has a large (regular) local Gowers norm. It turns out that such a reduction is always possible by a simple averaging argument,

[9] We write $(B_j)_{j \in X}$ for $\prod_{j \in X} B_j$ in subscripts.

together with an application of the Gowers-Cauchy-Schwarz inequality to separate the translated functions arising from such a process.

PROPOSITION 8.11. *Let $\eta \in (0,1]$ be a parameter. Suppose that B_0, B_1, B_2 are regular d-dimensional Bohr sets such that $B_1, B_2 \leqslant_\rho B_0$, and consider two other Bohr sets $\widetilde{B}_1 \leqslant_{\widetilde{\rho}} B_1$ and $\widetilde{B}_2 \leqslant_{\widetilde{\rho}} B_2$, where $\rho, \widetilde{\rho} \leqslant c\eta^4/d$. Then for $f : \mathbb{Z}_M \to [-1,1]$ and $a, b \in \Xi$,*

$$\mathbb{E}_{u_0 \in B_0} \|f(u_0 + \cdot)\|^4_{\boxtimes_{a,b}(B_1 \times B_2)} \geqslant \eta^4 \Rightarrow \mathbb{E}_{u_0 \in B_0} \|f(u_0 + ab\,\cdot)\|^4_{\boxtimes(\widetilde{B}_1 \times \widetilde{B}_2)} \geqslant \eta^4/2$$

PROOF. Unfolding the definition of the twisted U^2 norm, we have

$$\eta^4 \leqslant \mathbb{E}_{u_0 \in B_0} \mathbb{E}_{x^{(0)}, x^{(1)} \in B_1 \times B_2} \prod_{\omega \in \{0,1\}^2} f(u_0 + a x_1^{(\omega_1)} + b x_2^{(\omega_2)}).$$

By regularity, we now duplicate the variables $x_1^{(\varepsilon)}$ into $x_1^{(\varepsilon)} + b y_1^{(\varepsilon)}$ with $y_1^{(\varepsilon)} \in \widetilde{B}_1$, and the variables $x_2^{(\varepsilon)}$ into $x_2^{(\varepsilon)} + a y_2^{(\varepsilon)}$ with $y_2^{(\varepsilon)} \in \widetilde{B}_2$, so that

$$\begin{aligned}\eta^4 - O(\tilde{\rho} d) &\leqslant \mathbb{E}_{u_0 \in B_0} \mathbb{E}_{x^{(0)}, x^{(1)} \in B_1 \times B_2} \mathbb{E}_{y^{(0)}, y^{(1)} \in \widetilde{B}_1 \times \widetilde{B}_2} \\ &\qquad \prod_{\omega \in \{0,1\}^2} f\Big(u_0 + a x_1^{(\omega_1)} + b x_2^{(\omega_2)} + ab(y_1^{(\omega_1)} + y_2^{(\omega_2)})\Big) \\ &= \mathbb{E}_{u_0 \in B_0} \mathbb{E}_{x^{(0)}, x^{(1)} \in B_1 \times B_2} \langle (f(u_0 + a x_1^{(\omega_1)} + b x_2^{(\omega_2)} + abS))_\omega \rangle_{\square(\widetilde{B}_1 \times \widetilde{B}_2)},\end{aligned}$$

where $S : \widetilde{B}_1 \times \widetilde{B}_2 \to \mathbb{Z}_M$ is defined by $S(u_1, u_2) = u_1 + u_2$. Applying successively the Gowers-Cauchy-Schwarz inequality (8.2) and Hölder's inequality, we obtain

$$\begin{aligned}c\eta^{16} &\leqslant \Big(\mathbb{E}_{u_0 \in B_0} \mathbb{E}_{x^{(0)}, x^{(1)} \in B_1 \times B_2} \prod_{\omega \in \{0,1\}^2} \|f(u_0 + a x_1^{(\omega_1)} + b x_2^{(\omega_2)} + abS)\|_{\square(\widetilde{B}_1 \times \widetilde{B}_2)}\Big)^4 \\ &\leqslant \prod_{\omega \in \{0,1\}^2} \mathbb{E}_{u_0 \in B_0} \mathbb{E}_{x^{(0)}, x^{(1)} \in B_1 \times B_2} \|f(u_0 + a x_1^{(\omega_1)} + b x_2^{(\omega_2)} + abS)\|^4_{\square(\widetilde{B}_1 \times \widetilde{B}_2)}.\end{aligned}$$

By the pigeonhole principle, we may therefore find $\omega \in \{0,1\}^2$ such that

$$\begin{aligned}c\eta^4 &\leqslant \mathbb{E}_{u_0 \in B_0} \mathbb{E}_{x^{(0)}, x^{(1)} \in B_1 \times B_2} \|f(u_0 + a x_1^{(\omega_1)} + b x_2^{(\omega_2)} + abS)\|^4_{\square(\widetilde{B}_1 \times \widetilde{B}_2)} \\ &= \mathbb{E}_{u_0 \in B_0} \|f(u_0 + abS)\|^4_{\square(\widetilde{B}_1 \times \widetilde{B}_2)} + O(\rho d),\end{aligned}$$

where we have used regularity in the variable u_0 in the last step. The proposition follows from recalling Definition 8.8. □

At this point, we have reduced to a situation where we may apply Shao's local inverse U^2 theorem [**91**, Theorem 3.2 and Lemma 5.1], quoted below, to obtain a density increment. The presence of a coefficient $m = ab$ calls for a minor variant[10] of that result, which can however be effortlessly extracted out of Shao's argument: we omit the proof. Note also that in the proposition below, we consider Bohr sets of $\mathbb{Z}_M$ as sets of integers via the pullback of $\pi : [-M/2, M/2]_{\mathbb{Z}} \xrightarrow{\sim} \mathbb{Z}_M$.

PROPOSITION 8.12 (Local inverse U^2 theorem [**91**]). *Let $\eta \in (0, \frac{1}{2}]$ and $m \in \Xi \cdot \Xi$ be parameters. Suppose that B_0, B_1, B_2 are regular d-dimensional Bohr sets such that $B_1 \leqslant_\rho B_0$ and $B_2 \leqslant_\rho B_1$, where $\rho \leqslant c\eta^{12}/d$. Suppose also that $f : \mathbb{Z}_M \to [-1, 1]$ is such that $\mathbb{E}_{B_0} f = 0$ and*

$$\mathbb{E}_{u_0 \in B_0} \|f(u_0 + m\,\cdot\,)\|^4_{\boxtimes(B_1 \times B_2)} \gg \eta^4.$$

Then there exists $u \in \mathbb{Z}$ and a regular Bohr set B_3 such that $u + mB_3 \subset B_0$ in $\mathbb{Z}$, and

$$d_3 \leqslant d + 1, \qquad \delta_3 \geqslant (\eta/d)^{O(1)}\delta_1, \qquad \mathbb{E}_{u+mB_3} f \geqslant c\eta^{12}.$$

We are now ready to combine the previous propositions into our main density-increment statement, which we then iterate to obtain Proposition 8.1.

PROPOSITION 8.13 (Main iterative proposition). *Suppose that A is a subset of density $\alpha \in (0, \frac{1}{2}]$ of a regular d-dimensional Bohr set B contained in $[-N, N]$. Then either*

[10] Note also that Bohr sets on $\mathbb{Z}$ are used in that reference, however this is only a cosmetic difference. We actually quote a slightly weaker, but simpler, one-case consequence of Shao's result to fluidify our argument.

(i) *(Many φ-configurations) we have*

$$\#\{x \in [-N,N]^{q+1} : \varphi(x) \in A^t\} \geqslant (\alpha\delta/d)^{O(d)} N^{q+1},$$

(ii) *(Density increment) or there exists $u \in \mathbb{Z}$, $m \in \mathbb{N}$ and a regular Bohr set B' such that $u + mB' \subset B$ in $\mathbb{Z}$ and, writing $\alpha' = |A \cap (u+mB')|/|B'|$,*

$$\alpha' \geqslant (1 + c\alpha^{12t-1})\alpha, \qquad d' \leqslant d+1, \qquad \delta' \geqslant (\alpha/d)^{O(1)}\delta.$$

PROOF. Write $\eta = \alpha^t$ and choose $\rho = c\eta^{12}/d$. Let $B_0 = B$, and choose regular Bohr sets $B_1, \dots, B_q$ with $B_i = B_{i-1|\rho_i}$ and $\rho_i \in [\rho/2, \rho]$, so as to apply Proposition 8.9. Since $B_i \subset [-N,N]$ and $M > 2\|\varphi\|N$, for any $x \in B_0 \times \cdots \times B_q$, $\varphi(x)$ belongs to A^t modulo M if and only if it does in $\mathbb{Z}$. Therefore, if we are in the first case of Proposition 8.9, we have

$$\#\{x \in [-N,N]^{q+1} : \varphi(x) \in A^t\} \geqslant c\alpha^t |B_0| \dots |B_q| \geqslant (\alpha\delta/d)^{O(d)} M^{q+1}. \tag{8.4}$$

In the second case, we deduce, by Proposition 8.10, that there exist $i \in [t]$, $1 \leqslant k < \ell \leqslant q$ and twists $a, b \in \Xi_i$ such that, for $f_A = 1_A - \alpha 1_{B_0}$,

$$\mathbb{E}_{u_0 \in B_0} \|f_A(u_0 + \cdot)\|^4_{\boxtimes_{a,b}(B_k \times B_\ell)} \gg \eta^4.$$

Via Proposition 8.11, we may assume instead that

$$\mathbb{E}_{u_0 \in B_0} \|f_A(u_0 + ab\,\cdot)\|^4_{\boxtimes(\widetilde{B}_k \times \widetilde{B}_\ell)} \gg \eta^4$$

for regular dilates $\widetilde{B}_k = B_{k|\rho_k}$ and $\widetilde{B}_\ell = B_{\ell|\rho_\ell}$ with $\rho_k, \rho_\ell \in [\rho/2, \rho]$; note that we have $\widetilde{B}_k \leqslant_{2\rho} \widetilde{B}_\ell$. Finally, an application of Proposition 8.12 to f_A yields a density increment of the desired shape. □

Proof of Proposition 8.1. As stated at the beginning of this section, we use a parametrization $\varphi : \mathbb{Z}^{q+1} \twoheadrightarrow \mathbb{Z}^t \cap \mathrm{Ker}_{\mathbb{Q}}(V)$, so that $\mathrm{rk}(\varphi) = \dim(\mathrm{Ker}_{\mathbb{Q}} V) = t - r$. We embed $[-N,N]$ in a regular Bohr set $B^{(0)} := B(\{1\}, \frac{c}{D})$ of $\mathbb{Z}_M$, where $c \in [1,2]$

and $M = DN$. The set $A^{(0)} := A$ then has density $\gg \alpha$ in $B^{(0)}$. We now construct iteratively a sequence of regular Bohr sets $B^{(i)}$ of dimension d_i and radius δ_i contained in $[-N, N]$, and a sequence of subsets A_i of $B^{(i)}$ of density α_i; we also view A_i as subsets of $\mathbb{Z}$ via the pullback of $\pi : [-M/2, M/2]_{\mathbb{Z}} \xrightarrow{\sim} \mathbb{Z}_M$. At each step we apply Proposition 8.13 to the set A_i, and in the second case of that proposition we define A_{i+1} in $\mathbb{Z}$ by

$$A_i \cap (u_{i+1} + m_{i+1}B_{i+1}) = u_{i+1} + m_{i+1}A_{i+1}.$$

Writing $S_\varphi(Y) = \#\{x \in [-N, N]^{q+1} : \varphi(x) \in Y^t\}$ for a set of integers Y, it follows from the linearity and the presence of a shift variable in φ that $S_\varphi(A) \geqslant S_\varphi(A_i)$ for every i.

From $\alpha_{i+1} \geqslant (1 + c\alpha_i^{12t-1})\alpha_i$ and a familiar geometric series summation [**27**, Chapter 6], we deduce that the algorithm runs for at most $O(\alpha^{-12t+1})$ steps. Iterating the dimension and radius bounds, we also deduce that $d_i \ll \alpha^{-12t+1}$ and $\delta_i \geqslant \exp[-C\alpha^{-12t+1}\log\alpha^{-1}]$. Bounding crudely $\alpha^2 \log\alpha^{-1} \ll 1$, we have therefore, in the first case of Proposition 8.13,

$$\#\{x \in [-N, N]^{q+1} : \varphi(x) \in A^t\} \geqslant \exp\left[-C\alpha^{-24t}\right] \cdot N^{q+1}. \tag{8.5}$$

Since φ has rank $t - r$, for each $y \in [N]^t$, we have the multiplicity bound

$$\#\{x \in [-N, N]^{q+1} : \varphi(x) = y\} \ll N^{(q+1)-(t-r)}.$$

Summing over values $y = \varphi(x)$ in (8.5), we have therefore

$$\#\{y \in A^t : Vy = 0\} \geqslant \exp\left[-C\alpha^{-24t}\right] \cdot N^{t-r}.$$

□

9. Appendix: On Roth's matrix conditions

In this appendix we discuss in more detail the notion of complexity one, and we compare it with an earlier class of systems of equations considered by Roth [70]. Here we view linear forms on $\mathbb{Z}^d$ for $d \geqslant 1$ as linear forms on $\mathbb{Q}^d$, and we carry out all further linear algebra manipulations with respect to the base field $\mathbb{Q}$. For two vectors $u, v \in \mathbb{Q}^d$, we also let $u \cdot v$ denote the canonical scalar product of u and v, and we write $A^\perp$ for the orthogonal of a subset A of $\mathbb{Q}^d$. We now state Roth's matrix conditions [70], which we term, somewhat anachronously, "Roth complexity".

Definition 9.1 (Roth complexity). *Let $V = [C_1 \cdots C_t] \in \mathcal{M}_{r\times t}(\mathbb{Z})$. We say that V has Roth complexity at $i \in [t]$ when there exists a partition $[t] \smallsetminus \{i\} = Y_1 \sqcup Y_2 \sqcup Z$ with $|Y_1| = |Y_2| = r$ such that, for every $k \in \{1, 2\}$, the columns $(C_j, j \in X_k)$ are linearly independent. We say that V has Roth complexity when there exists a set $J \subset [t]$ with $|J| = r$ such that the columns $(C_j, j \in J)$ are linearly independent, and such that V has Roth complexity at every $i \in J$.*

Roth [70] has shown that a translation-invariant system of equations of the above type is non-trivially solvable in any subset of $[N]$ of density at least $C(\log\log N)^{-1/r^2}$. Definition 9.1 is motivated by Fourier analysis: if $C_1, \ldots, C_t$ are the columns of V and A is a subset of $\mathbb{Z}_M$ of density α, the normalized count of solutions $y \in A^t$ to $Vy = 0$ has a Fourier expression

$$\mathbb{E}_{y\in\mathbb{Z}_M^t : Vy=0} A(y_1)\cdots A(y_t) = \alpha^t + \sum_{u\in\mathbb{Z}_M^t\smallsetminus\{0\}} \widehat{A}(C_1\cdot u)\cdots\widehat{A}(C_t\cdot u).$$

For every $u \neq 0$, we may find $i \in J$ such that $C_i \cdot u \neq 0$, where J is the set from Definition 9.1. The assumption of Roth complexity then ensures, via an L^∞–L^2–L^2 bound, that the sum over $u \neq 0$ is bounded by $\sup_{r\neq 0} |\widehat{A}(r)|$, and Roth's proof [70] then follows the nowadays standard strategy of density increment on arithmetic progressions. This argument has been revisited recently by Liu, Spencer

and Zhao [61,62], who extended it to the setting of function fields and finite abelian groups. We now compare the notion of Roth complexity to that of complexity at most one from Section 4, whose definition we recall now.

Definition 9.2 (Complexity zero/one). *Consider a system of linear forms $\psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ with $t \geqslant 3$. We say that ψ has complexity at most one at $i \in [t]$ when there exists a partition $[t] \smallsetminus \{i\} = X_1 \sqcup X_2$ into non-empty sets such that*

$$\psi_i \notin \langle \psi_j, j \in X_k \rangle \quad \forall k \in \{1, 2\}.$$

Furthemore, we say that ψ has complexity zero at $i \in [t]$ when $\psi_i \notin \langle \psi_j, j \neq i \rangle$.

Recall also that the complexity of a matrix $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ at a position $i \in [t]$ is defined to be that of any linear surjection $\psi : \mathbb{Q}^d \twoheadrightarrow \mathrm{Ker}_{\mathbb{Q}}(V)$, and we have verified in Proposition 4.7 that this constitutes a valid definition. We now develop a more convenient criterion in the case of complexity zero or one.

Proposition 9.3 (Complexity zero/one criterion). *Let $V = [C_1 \cdots C_t] \in \mathcal{M}_{r\times t}(\mathbb{Z})$ with $t \geqslant 3$. Then V has complexity at most one at $i \in [t]$ if and only if there exists a partition $[t] \smallsetminus \{i\} = X_1 \sqcup X_2$ into non-empty sets such that*

$$C_i \in \langle C_j, j \in X_k \rangle \quad \forall k \in \{1, 2\}.$$

Furthermore, V has complexity zero at $i \in [t]$ if and only if $C_i = 0$.

Proof. Denote by $L_1, \dots, L_r \in \mathcal{M}_{1\times t}(\mathbb{Z})$ the lines of V, and consider a surjection $\psi : \mathbb{Q}^d \twoheadrightarrow \mathrm{Ker}_{\mathbb{Q}}(V)$ and an indice $i \in [t]$. We start with the proof of the complexity-one criterion, and we fix a partition $[t] \smallsetminus \{i\} = X_1 \sqcup X_2$ into non-empty sets. As in the proof of Proposition 4.7, we have

$$\psi_i \in \langle \psi_j, j \in X_k \rangle \Leftrightarrow (e_i \oplus_{j\in X_k} \mathbb{Q}e_j) \cap \langle {}^{\mathrm{t}}L_1, \dots, {}^{\mathrm{t}}L_r \rangle \neq \varnothing, \tag{9.1}$$

where $(e_i)_{1\leqslant i\leqslant t}$ is the canonical basis of $\mathbb{Q}^t$. We next show that

$$(e_i \oplus_{j\in X_1} \mathbb{Q}e_j) \cap \langle {}^{\mathrm{t}}L_1, \dots, {}^{\mathrm{t}}L_r\rangle \neq \varnothing \Leftrightarrow C_i \notin \langle C_j, j \in X_2\rangle; \tag{9.2}$$

an analogous statement also holds with the roles of X_1 and X_2 reversed. By orthogonality, the left-hand side of (9.2) is equivalent to the existence of $\mu \in \mathbb{Q}^r$ such that

$$\textstyle\sum_{j=1}^r \mu_j \, {}^{\mathrm{t}}L_j \cdot e_i = 1 \quad \text{and} \quad \sum_{j=1}^r \mu_j \, {}^{\mathrm{t}}L_j \cdot e_m = 0 \quad \forall m \in X_2.$$

Since ${}^{\mathrm{t}}L_j \cdot e_m$ is the j-th element of the column C_m, this is equivalent to

$$\mu \cdot C_i = 1 \quad \text{and} \quad \mu \cdot C_m = 0 \quad \forall m \in X_2.$$

Upto renormalizing, the existence of $\mu \in \mathbb{Q}^r$ satisfying the above is equivalent to

$$\exists \mu \in \langle C_m, m \in X_2\rangle^{\perp} : \mu \cdot C_i \neq 0 \quad \Leftrightarrow \quad C_i \notin \langle C_m, m \in X_2\rangle^{\perp\perp},$$

and by biorthogonality this concludes the proof of (9.2). The complexity-one criterion then follows by considering the contrapositives of (9.1) and (9.2).

To obtain the complexity-zero criterion, it is enough to observe that one has, by the same arguments as before,

$$\begin{aligned}\psi_i \in \langle \psi_j, j \neq i\rangle &\Leftrightarrow \left(e_i + \textstyle\sum_{j\neq i} \mathbb{Q}e_j\right) \cap \langle {}^{\mathrm{t}}L_1, \dots, {}^{\mathrm{t}}L_r\rangle \neq \varnothing \\ &\Leftrightarrow \exists \mu \in \mathbb{Q}^r : \textstyle\sum_{j=1}^r \mu_j \, {}^{\mathrm{t}}L_j \cdot e_i = 1 \\ &\Leftrightarrow \exists \mu \in \mathbb{Q}^r : \mu \cdot C_i \neq 0,\end{aligned}$$

and this last condition is satisfied if and only if C_i is non-zero. □

COROLLARY 9.4. *Let $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ and $i \in [t]$. If V has Roth complexity at i, it has complexity at most one at i.*

PROOF. We have in particular $t \geqslant 2r+1 \geqslant 3$. Partitioning $[t] \setminus \{i\} = Y_1 \sqcup Y_2 \sqcup Z$ as in Definition 9.1, and letting $X_1 = Y_1$ and $X_2 = Y_2 \sqcup Z$, we see by simple linear algebra that $C_i \in \langle C_j, j \in X_k \rangle$ for every $k \in \{1, 2\}$. □

This shows that a slightly stronger notion of Roth complexity, where one assumes Roth complexity at *every* position i, is subsumed by the notion of complexity one. We have not been able to determine definitively whether matrices of Roth complexity do have complexity one. Since these definitions of complexity arise from quite different underlying techniques to bound averages over linear patterns, it may well be that they correspond to different classes of systems of equations. The most we can say is that systems of Roth complexity have finite complexity, by the following argument. If $V \in \mathcal{M}_{r\times t}(\mathbb{Q})$ with $t \geqslant 2r+1$ has infinite complexity, its row space contains a non-zero vector with at most two non-zero entries (by the usual orthogonality argument). Up to multiplication by an invertible matrix, we may assume this vector to be a line of V, and one of its non-zero entries must then belong to a column from the set J of r invertible columns from Definition 9.1. But it is then impossible to form two invertible matrices when that column is excluded, since one of them is bound to contain a zero line.

10. Appendix: Consequences of higher-complexity theorems

In this section we record certain results on translation-invariant equations which follow at once from Gowers' proof [**20**] of Szemerédi's theorem [**95**], and the extension of the latter to the primes by Green and Tao [**36**]. We are very grateful to Pablo Candela for showing us the arguments below.

THEOREM 10.1 (Gowers). *Suppose that $V \in \mathcal{M}_{r\times t}(\mathbb{Z})$ is a translation-invariant matrix of rank r and finite complexity, and A is a subset of $[N]$ of density at least*

$$C(\log\log N)^{-c_t},$$

where $c_t = 2^{-2^{t+9}}$ *and* $C > 0$ *is a constant depending at most on* r, t, V*. Then there exists a solution* $\mathbf{y} \in A^t$ *to* $V\mathbf{y} = 0$ *with distinct coordinates.*

PROOF. By Proposition 4.8, we may consider a linear surjection $\varphi : \mathbb{Z}^{d+1} \twoheadrightarrow \mathbb{Z}^t \cap \operatorname{Ker} V$ of the form $\varphi(x_0, x) = x_0\mathbf{1} + \psi(x)$, where $\psi = (\psi_1, \dots, \psi_t)$ has finite complexity, so that no two forms ψ_i, ψ_j with $i \neq j$ are linearly dependent. Therefore, each equation $\psi_i = \psi_j$ defines a hyperplane of $\mathbb{Q}^d$, and it is then easy to find an integer $u \in \mathbb{Z}^d$ such that the values $c_i = \psi_i(u), i \in [t]$ are all distinct. But then, by the same argument as for arithmetic progressions, the system

$$\Upsilon(y, d) = (y + c_1 d, \dots, y + c_t d) \tag{10.1}$$

is controlled by the Gowers U^{t-1} norm. By Gowers' density-increment strategy [**20**], it follows that A^t contains a distinct-coordinates configuration $\Upsilon(y, d) = \varphi(y, du)$. □

THEOREM 10.2 (Green-Tao). *Suppose that* V *is a translation-invariant matrix of finite complexity, and* A *is a subset of the primes of positive upper density. Then there exists a solution* $\mathbf{y} \in A^t$ *to* $V\mathbf{y} = 0$ *with distinct coordinates.*

PROOF. The beginning of the proof is identical to that of Theorem 10.1, so that we are led to identifying distinct-coordinates configurations of the form (10.1) in A^t. Since this system has finite complexity, the result follows from [**36**], using Theorem 10.1 in place of Szemerédi's theorem there, and the finite-complexity generalized Von Neumann theorem from [**39**, Appendix C] in place of [**36**, Proposition 5.3]. One should also follow the remarks in [**36**, Section 11] on how to adapt the arguments to a dense subset of the primes instead of the set of all primes. □

Bibliographie

1. M. Bateman and N. H. Katz, *New bounds on cap sets*, J. Amer. Math. Soc. **25** (2012), no. 2, 585–613.
2. T. F. Bloom, *Translation invariant equations and the method of Sanders*, Bull. Lond. Math. Soc. **44** (2012), no. 5, 1050–1067.
3. J. Bourgain, *On $\Lambda(p)$-subsets of squares*, Israel J. Math. **67** (1989), no. 3, 291–311.
4. ———, *On arithmetic progressions in sums of sets of integers*, A tribute to Paul Erdős, Cambridge Univ. Press, Cambridge, 1990, pp. 105–109.
5. ———, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), no. 5, 968–984.
6. ———, *Roth's theorem on progressions revisited*, J. Anal. Math. **104** (2008), 155–192.
7. M.-C. Chang, *A polynomial bound in Freiman's theorem*, Duke Math. J. **113** (2002), no. 3, 399–419.
8. D. Conlon, J. Fox, and Y. Zhao, *The Green-Tao theorem, an exposition*, Preprint (2014), http://arxiv.org/abs/1403.2957.
9. E. Croot, I. Łaba, and O. Sisask, *Arithmetic progressions in sumsets and L^p-almost-periodicity*, Combin. Probab. Comput. **22** (2013), no. 3, 351–365.
10. E. Croot, I. Z. Ruzsa, and T. Schoen, *Arithmetic progressions in sparse sumsets*, Combinatorial number theory, de Gruyter, Berlin, 2007, pp. 157–164.
11. E. Croot and O. Sisask, *A probabilistic technique for finding almost-periods of convolutions*, Geom. Funct. Anal. **20** (2010), no. 6, 1367–1396.
12. Z. Cui, H. Li, and B. Xue, *Long arithmetic progressions in $A + A + A$ with A a prime subset*, J. Number Theory **132** (2012), no. 7, 1572–1582.
13. H. Davenport, *Multiplicative number theory*, third ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000.
14. H. G. Diamond and H. Halberstam, *A higher-dimensional sieve method*, Cambridge Tracts in Mathematics, vol. 177, Cambridge University Press, Cambridge, 2008.
15. J. Dousse, *On a generalisation of Roth's theorem for arithmetic progressions and applications to sum-free subsets*, Math. Proc. Cambridge Philos. Soc. **155** (2013), no. 2, 331–341.

16. P. Erdös and P. Turán, *On some sequences of integers*, J. London Math. Soc. **S1-11** (1936), no. 4, 261.
17. G. A. Freĭman, *Foundations of a structural theory of set addition*, American Mathematical Society, Providence, R. I., 1973, Translations of Mathematical Monographs, Vol 37.
18. G. A. Freiman, H. Halberstam, and I. Z. Ruzsa, *Integer sum sets containing long arithmetic progressions*, J. London Math. Soc. (2) **46** (1992), no. 2, 193–201.
19. D. A. Goldston, J. Pintz, and C. Y. Yıldırım, *Primes in tuples. I*, Ann. of Math. (2) **170** (2009), no. 2, 819–862.
20. W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588.
21. ———, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), no. 3, 897–946.
22. ———, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem*, Bull. Lond. Math. Soc. **42** (2010), no. 4, 573–606.
23. W. T. Gowers and J. Wolf, *The true complexity of a system of linear equations*, Proc. Lond. Math. Soc. (3) **100** (2010), no. 1, 155–176.
24. ———, *Linear forms and higher-degree uniformity for functions on* $\mathbb{F}_p^n$, Geom. Funct. Anal. **21** (2011), no. 1, 36–69.
25. ———, *Linear forms and quadratic uniformity for functions on* $\mathbb{F}_p^n$, Mathematika **57** (2011), no. 2, 215–237.
26. ———, *Linear forms and quadratic uniformity for functions on* $\mathbb{Z}_N$, J. Anal. Math. **115** (2011), 121–186.
27. A. Granville and B. Green, *Additive combinatorics*, Upcoming book (2014).
28. B. Green, *On triples in arithmetic progressions*, Expository note (1999), http://people.maths.ox.ac.uk/greenbj/papers/bourgain-roth.pdf.
29. ———, *Arithmetic progressions in sumsets*, Geom. Funct. Anal. **12** (2002), no. 3, 584–597.
30. ———, *Roth's theorem in the primes*, Ann. of Math. (2) **161** (2005), no. 3, 1609–1636.
31. B. Green and S. Konyagin, *On the Littlewood problem modulo a prime*, Canad. J. Math. **61** (2009), no. 1, 141–164.
32. B. Green and I. Z. Ruzsa, *Freiman's theorem in an arbitrary abelian group*, J. Lond. Math. Soc. (2) **75** (2007), no. 1, 163–175.
33. B. Green and T. Sanders, *A quantitative version of the idempotent theorem in harmonic analysis*, Ann. of Math. (2) **168** (2008), no. 3, 1025–1054.

34. B. Green and T. Tao, *Restriction theory of the Selberg sieve, with applications*, J. Théor. Nombres Bordeaux **18** (2006), no. 1, 147–182.

35. ———, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edinb. Math. Soc. (2) **51** (2008), no. 1, 73–153.

36. ———, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547.

37. ———, *An arithmetic regularity lemma, an associated counting lemma, and applications*, An irregular mind, Bolyai Soc. Math. Stud., vol. 21, 2010, pp. 261–334.

38. ———, *An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm*, Math. Proc. Cambridge Philos. Soc. **149** (2010), no. 1, 1–19.

39. ———, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850.

40. ———, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. (2) **175** (2012), no. 2, 465–540.

41. ———, *On the quantitative distribution of polynomial nilsequences—erratum*, Ann. of Math. (2) **179** (2014), no. 3, 1175–1183.

42. B. Green, T. Tao, and T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), no. 2, 1231–1372.

43. H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press, London-New York, 1974, London Mathematical Society Monographs, No. 4.

44. M. Hamel, N. Lyall, K. Thompson, and N. Walters, *Arithmetic structure in sparse difference sets*, J. Number Theory **130** (2010), no. 7, 1581–1589.

45. H. Hatami, *Fourier analysis of finite abelian groups*, Lecture note (2011), http://cs.mcgill.ca/~hatami/comp760-2014/lectures.pdf.

46. H. Hatami, P. Hatami, and S. Lovett, *General systems of linear forms ; equidistribution and true complexity*, Preprint (2014), http://arxiv.org/abs/1403.7703.

47. H. Hatami and S. Lovett, *Higher-order Fourier analysis of $\mathbb{F}_p^n$ and the complexity of systems of linear forms*, Geom. Funct. Anal. **21** (2011), no. 6, 1331–1357.

48. D. R. Heath-Brown, *Integer sets containing no arithmetic progressions*, J. London Math. Soc. (2) **35** (1987), no. 3, 385–394.

49. N. Hegyvári, F. Hennecart, and A. Plagne, *A proof of two Erdős' conjectures on restricted addition and further results*, J. Reine Angew. Math. **560** (2003), 199–220.

50. H. A. Helfgott and A. de Roton, *Improving Roth's theorem in the primes*, Int. Math. Res. Not. IMRN (2011), no. 4, 767–783.

51. K. Henriot, *Arithmetic progressions in sets of small doubling*, Preprint (2013), http://arxiv.org/abs/1308.5248.
52. _____, *Bourgain's bounds for Roth's theorem*, Expository note (2013), http://dms.umontreal.ca/~henriot/bourgainroth.pdf.
53. _____, *Notes on the Croot-Sisask lemma*, Expository note (2013), http://dms.umontreal.ca/~henriot/almostp.pdf.
54. _____, *On arithmetic progressions in $A+B+C$*, Int. Math. Res. Not. (2013), Published online at http://imrn.oxfordjournals.org/content/early/2013/06/11/imrn.rnt121.abstract.
55. S. Johnson, *Saddle-point integration of C^∞ bump functions*, Expository note (2006), http://math.mit.edu/~stevenj/bump-saddle.pdf.
56. N. H. Katz and P. Koester, *On additive doubling and energy*, SIAM J. Discrete Math. **24** (2010), no. 4, 1684–1693.
57. Y. Katznelson, *An introduction to harmonic analysis*, third ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2004.
58. I. Łaba, *From harmonic analysis to arithmetic combinatorics*, Bull. Amer. Math. Soc. (N.S.) **45** (2008), no. 1, 77–115.
59. T. H. Le and J. Wolf, *Polynomial configurations in the primes*, To appear in Int. Math. Res. Not. (2013).
60. V. F. Lev, *Progression-free sets in finite abelian groups*, J. Number Theory **104** (2004), no. 1, 162–169.
61. Y.-R. Liu, C. V. Spencer, and X. Zhao, *Roth's theorem on systems of linear forms in function fields*, Acta Arith. **142** (2010), no. 4, 377–386.
62. _____, *A generalization of Meshulam's theorem on subsets of finite abelian groups with no 3-term arithmetic progression (II)*, European J. Combin. **32** (2011), no. 2, 258–264.
63. S. Lovett, *An exposition of Sanders' quasi-polynomial Freiman-Ruzsa theorem*, Expository note (2012), http://eccc.hpi-web.de/report/2012/029/download.
64. N. Lyall, *Behrend's example*, Expository note (2005), http://www.math.uga.edu/%7Elyall/REU/Behrend.pdf.
65. E. Naslund, *On improving Roth's theorem in the primes*, To appear in Mathematika (2014), First arxiv version : http://arxiv.org/abs/1302.2299v1, Second arxiv version : http://arxiv.org/abs/1302.2299.

66. G. Petridis, *New proofs of Plünnecke-type estimates for product sets in groups*, Combinatorica **32** (2012), no. 6, 721–733.

67. O. Ramaré, *On Šnirel′man's constant*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **22** (1995), no. 4, 645–706.

68. O. Ramaré and I. Z. Ruzsa, *Additive properties of dense subsets of sifted sequences*, J. Théor. Nombres Bordeaux **13** (2001), no. 2, 559–581.

69. K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.

70. ———, *On certain sets of integers. II*, J. London Math. Soc. **29** (1954), 20–26.

71. W. Rudin, *Trigonometric series with gaps*, J. Math. Mech. **9** (1960), 203–227.

72. ———, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987.

73. ———, *Fourier analysis on groups*, Wiley Classics Library, John Wiley & Sons, Inc., New York, 1990.

74. I. Z. Ruzsa, *Arithmetic progressions in sumsets*, Acta Arith. **60** (1991), no. 2, 191–202.

75. ———, *Arithmetical progressions and the number of sums*, Period. Math. Hungar. **25** (1992), no. 1, 105–111.

76. ———, *Solving a linear equation in a set of integers. I*, Acta Arith. **65** (1993), no. 3, 259–282.

77. ———, *Generalized arithmetical progressions and sumsets*, Acta Math. Hungar. **65** (1994), no. 4, 379–388.

78. T. Sanders, *Additive structures in sumsets*, Math. Proc. Cambridge Philos. Soc. **144** (2008), no. 2, 289–316.

79. ———, *Roth's theorem in* $\mathbb{Z}_4^n$, Anal. PDE **2** (2009), no. 2, 211–234.

80. ———, *Three-term arithmetic progressions and sumsets*, Proc. Edinb. Math. Soc. (2) **52** (2009), no. 1, 211–233.

81. ———, *On Roth's theorem on progressions*, Ann. of Math. (2) **174** (2011), no. 1, 619–636.

82. ———, *On certain other sets of integers*, J. Anal. Math. **116** (2012), 53–82.

83. ———, *On the Bogolyubov-Ruzsa lemma*, Anal. PDE **5** (2012), no. 3, 627–655. MR 2994508

84. ———, *The structure theory of set addition revisited*, Bull. Amer. Math. Soc. (N.S.) **50** (2013), no. 1, 93–127.

85. W. M. Schmidt, *Diophantine approximation*, Lecture Notes in Mathematics, vol. 785, Springer, Berlin, 1980.

86. T. Schoen, *The cardinality of restricted sumsets*, J. Number Theory **96** (2002), no. 1, 48–54.

87. ———, *Linear equations in* $\mathbb{Z}_p$, Bull. London Math. Soc. **37** (2005), no. 4, 495–501.

88. ———, *Near optimal bounds in Freiman's theorem*, Duke Math. J. **158** (2011), no. 1, 1–12.

89. ———, *Linear equations and sets of integers*, Acta Math. Hungar. **135** (2012), no. 3, 229–235.
90. T. Schoen and I. D. Shkredov, *Roth's theorem in many variables*, Preprint (2011), http://arxiv.org/abs/1106.1601.
91. X. Shao, *Finding linear patterns of complexity one*, To appear in Int. Math. Res. Not. IMRN (2013), http://arxiv.org/abs/1309.0644.
92. A. Shapira, *Behrend-type constructions for sets of linear equations*, Acta Arith. **122** (2006), no. 1, 17–33.
93. J. Solymosi, *Arithmetic progressions in sets with small sumsets*, Combin. Probab. Comput. **15** (2006), no. 4, 597–603.
94. Y. V. Stanchescu, *Planar sets containing no three collinear points and non-averaging sets of integers*, Discrete Math. **256** (2002), no. 1-2, 387–395.
95. E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.
96. ———, *Integer sets containing no arithmetic progressions*, Acta Math. Hungar. **56** (1990), no. 1-2, 155–158.
97. T. Tao, *Montgomery's uncertainty principle*, Blog post (2011), http://terrytao.wordpress.com/2011/12/31/montgomerys-uncertainty-principle/.
98. ———, *Notes on linear patterns*, Blog post (2010), http://terrytao.wordpress.com/2010/04/23/254b-notes-3-linear-patterns/#more-3708.
99. ———, *Higher order Fourier analysis*, Graduate Studies in Mathematics, vol. 142, American Mathematical Society, Providence, RI, 2012.
100. T. Tao and V. H. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2010.
101. T. Tao and T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), no. 2, 213–305.
102. ———, *Erratum to "The primes contain arbitrarily long polynomial progressions"*, Acta Math. **210** (2013), no. 2, 403–404.
103. G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge Studies in Advanced Mathematics, vol. 46, Cambridge University Press, Cambridge, 1995.
104. R. C. Vaughan, *The Hardy-Littlewood method*, second ed., Cambridge Tracts in Mathematics, vol. 125, Cambridge University Press, Cambridge, 1997.
105. C. Vinuesa, *Asymptotics for magic squares of primes*, Preprint (2012), http://arxiv.org/abs/1207.3936.